RELATIVITY & GRAVITATION

T. PERCY NUNN, M.A., D.Sc.


THE VNIVERSITY OF LIVERPOOL
FIAT LVX
HÆC OTIA
STVDIA FOVENT

RELATIVITY AND GRAVITATION

RELATIVITY AND GRAVITATION

AN ELEMENTARY TREATISE UPON EINSTEIN'S THEORY

BY

T. PERCY NUNN, M.A., D.Sc.

PROFESSOR OF EDUCATION IN THE UNIVERSITY OF LONDON

AUTHOR OF
"EDUCATION: ITS DATA AND FIRST PRINCIPLES"; "THE TEACHING OF ALGEBRA"; "THE AIM AND ACHIEVEMENTS OF SCIENTIFIC METHOD", ETC.

LONDON
UNIVERSITY OF LONDON PRESS, LTD.
17 WARWICK SQUARE, E.C.4
1923

Printed in Great Britain for the UNIVERSITY OF LONDON PRESS, LTD.,
by HAZELL, WATSON & VINEY, LD., London and Aylesbury.

PREFACE

Books upon the Theory of Relativity which are not philosophical in aim generally fall into one of two classes. They are either popular expositions intended for readers who have next to no mathematics, or else serious treatises presupposing in the student a considerable technical equipment. The present work seeks to fill a modest place between the two groups. Its level of difficulty may be indicated by saying that it should be well within the scope of anyone who has read mathematics up to, or nearly up to, the pass standard required for a B.Sc. degree, and that explanations are given of all theorems and processes which such a reader is not likely to have met with or may reasonably have forgotten. In addition, the demonstrations are set out with a fullness which would be tiresome to an expert mathematician, but may nevertheless be welcomed by those whom they are intended to assist. In all cases of doubt I have, in fact, assumed that my reader desires an explanation, instead of paying him the sometimes embarrassing compliment of assuming that he could do without it.

In a purely expository treatise there is little room for originality; but the present work contains a feature which is, I think, an innovation and will, I hope, prove a useful one. For the average student the great difficulty in the theory of relativity is the tensor calculus, which he is told he must master before he can enjoy with

Einstein the triumph of predicting the famous eclipse effect and of explaining the anomalous behaviour of the planet Mercury. There must be many who have set out with high hopes only to be turned back, baffled by this formidable obstacle. Now, in Einstein's classical memoir of 1916 there are indications of a route to the "crucial phenomena" which does not pass by way of the theory of tensors. It would be wholly in accord with the history of mathematical thought if it turned out that the great master himself first used this route and only afterwards laid down the "high priori" road to his wonderful discoveries. In any case, that is the procedure I have adopted. I bring the reader by the easy path to the results upon which the interest of the educated world centres, and develop the tensor calculus subsequently as a criterion by which the soundness of those results may be tested.

In conformity with my limited purpose I have not touched upon the problems of electro-magnetism and have been silent upon Einstein's cosmogonal speculations and Weyl's geometrical theories. For students with the necessary equipment there is a full discussion of these fascinating subjects in the masterly *Mathematical Theory of Relativity* which has come from Professor Eddington as these pages were being completed. Also I have dealt but slightly with the "restricted" theory, leaving the reader who seeks a fuller treatment to find it in Dr. L. Silberstein's or some other book.

I wish my own book to be regarded as an exposition of the elements written by a layman for other laymen who are, so to speak, a few lessons behind him. It is based upon a study of Einstein's own papers—which

Messrs. Methuen have now published in English—helped out by Professor Eddington's well-known Report, Professor Jean Becquerel's lucid French treatise, and the recently published *Theory of Relativity* of Professor Whitehead. From Professor Whitehead's book I have borrowed anything that would fit into my scheme, and I regret that I could not take more. I have been compelled to refer to it in the text only very briefly, but have ventured to express the opinion that it is a work of high moment and that its appearance raises issues of critical importance in the mathematico-philosophical discussions which the genius of Einstein and Minkowski set moving.

I have to thank Professor E. H. Neville and Mr. D. F. Taylor for valuable assistance in correcting the proofs of the book. To Professor Neville my obligation is indeed greater than any formal acknowledgment could discharge. He has scrutinised every sentence I wrote, and has placed his consummate mathematical scholarship most generously at my disposal. Any errors or imperfections that remain in the text must be attributed entirely to the author's obtuseness or incapacity; for he has had the best criticism a man could wish for.

T. P. NUNN.

LONDON.
April 1923.

CONTENTS

RELATIVITY AND GRAVITATION

CHAPTER I

ABSOLUTE AND RELATIVE MOTION

§ 1. LIKE many other great scientific ideas, the Principle of Relativity, with which the name of Albert Einstein is imperishably associated, is rooted in observations familiar to everyone. Those most germane to our purpose are simple observations concerning motion. Most of us have from a pier-head watched a steamer cast off and quietly recede, and at another time, being ourselves on board, have had the queer experience of seeing the pier apparently receding from the steamer. Why do we say " apparently" in the latter case and not in the former ? Partly because we know that the motion occurs at the *fiat* of the captain, who orders the engines to start but has no power to shift the pier ; but mainly because the pier runs out from the solid shore, backed by the streets of the town and with miles of *terra firma* behind it. For these reasons we think of the steamer as " really " moving and the motion of the pier as mere illusion.

Now, although this explanation would satisfy the unsophisticated, all educated people, since the days of Copernicus, recognize that it contains an important element of convention. The solid earth is no more than the steamer " really " at rest ; an observer on the sun would see it spinning like a fretful midge and swinging

ceaselessly round in its annual orbit. If he shifted his standpoint to a fixed star he might observe that the sun itself with its train of planets is heading for the constellation Hercules. And what more do we mean by calling the star "fixed" than that its motion, carried out in the remote depths of space, requires a long period to reveal itself to terrestrial observers? In fact, in this restless and turbulent world is there anything motionless in an absolute sense and not merely in relation to something else assumed by a convenient fiction to be at rest?

A partial answer to the question was given long ago. According to Newton's mechanics, it is at least possible to decide whether a given body is really or only apparently rotating; for if the rotation is real the parts of the body are subject to a "centrifugal force" which would be absent if it were merely relative. Thus if humanity had grown up under a canopy of clouds so thick as wholly to hide the heavens and wholly to obliterate the distinction between day and night, men of science (working by artificial light!) might still have noted the bulging round the equator, have invented the experiment of Foucault's pendulum, and have observed the apparent movement of a gyroscopic axis; and from these phenomena might have deduced the existence and rate of the earth's rotation.

But the possibility of distinguishing the real from the apparent does not, on Newton's principles, extend to non-rotatory motion. Even if "impressed force" could be measured by some means other than the acceleration it is supposed to cause, absolute and relative translation could not be discriminated; for if a body's motion is rectilinear and uniform, it may move with enormous

absolute speed without the help of any impressed force. And since, as a matter of fact, there is no generally applicable criterion of impressed force except the body's change of motion, and that change can be measured only by reference to another body, it is impossible to be sure whether even an accelerated body, at a given moment, is in absolute motion or at absolute rest. Thus, though rotating bodies give themselves away, bodies (apparently) at rest or in translation keep their secrets.

A scientific mind ought no doubt to be ready to give up the idea of absolute motion and to accept all motions as equally relative and at the same time equally real. But in the first place, there is the fact that Newton's principles do offer a criterion for absolute rotation, so that it seems anomalous that no criterion for absolute translation should be discoverable. And in the second place, scientific minds are not wholly free from ordinary human prejudices, and naturally shrink from the distressing idea that in the welter of the world's flux nothing is absolutely fixed. It is not surprising, then, that the hypothesis of an all-pervading ether, when it emerged in the early nineteenth century, should have been welcomed not only for the immense help it promised in the development of physical theory, but also because it offered some sort of refuge from universal relativity. If we confine attention to material bodies, it may be that all we can assert is that while some are in relative motion with regard to one another, others are relatively to one another at rest; but it may yet be that some are also motionless in the ether, and in that case may be considered as at rest, if not "absolutely", yet in a very special and exclusive sense.

It is true that this notion about the ether has not been held consistently. I do not refer to its vibrations, for a thing may vibrate and yet remain where it is. Some physical phenomena have, however, been explained by the supposition that the ether is dragged along with the matter immersed in it; and this idea is certainly contrary to the notion that all translation may be referred to the ether in an absolute or quasi-absolute sense. But the whole amount of the ether involved, hypothetically, in these currents is "only a little one", and the movements, together with those of the material bodies, are still referred to the great inter-stellar ocean which transmits luminiferous and electro-magnetic tremors, but remains, as a whole, eternally in the same place. Moreover, the later developments of ether-theory tended to deprive it even of vibratory motion and to make it completely "stationary". Thus it is substantially true that in modern physics the ether came to be regarded as a universal "system of reference" for the motion of all material bodies; and although it might be denied that this is the same thing as supposing it to be absolutely at rest, the distinction between the two notions is a rather thin one.

§ 2. The view taken of the ether during the nineteenth century made it highly desirable to prove and measure movement through it in at least one instance. The research first carried out in 1881 by the American physicist Michelson, and repeated later with increasing accuracy by Michelson and Morley and by Morley and Miller, was designed to test the possibility of doing so. Their famous experiment has been often described and need not be described again here; it will suffice to remind

the reader of its essential point and of its outcome. Suppose a pulse of light to be emitted from a point O which is at rest in the ether; then the wave-front will expand in all directions from O with the speed c, c being the universal velocity with which disturbances are propagated through the ether. But if the pulse is emitted at a time when O is moving in a direction OA with constant speed v with reference to the ether, then the wave-front will separate from O along OA with the velocity $c - v$, and this will be, from the standpoint of an observer moving with O, the measured velocity of light in the direction OA.* Now let O be a point on the earth. Then since the earth boxes the compass annually in its voyage round the sun, it must sometimes be moving relatively to the ether, even when allowance is made for the drag of the solar system towards Hercules and for a possible wider drift of the whole visible universe. Its orbital speed is about 30 km. per sec., and the effect caused by this should have been measurable even in Michelson's earliest experiment. But no such effect was observed, and the repetitions of the experiment have proved that it cannot be more than one two-hundredth of the amount predicted by theory. In other words, it is now known that the velocity of light, to within one part in 10^{10}, is entirely independent of the motion of its source.

The obvious inference from this striking result has been generally accepted by physicists.† Even if we retain the hypothesis of the ether as the all-pervading

* The reader is aware that in practice it is possible to measure only the *average* speed of light during the return journey from O to some point A and back.

† It has been confirmed by experiments of an entirely different character due to Rayleigh, to Brace, and to Trouton and Noble.

vehicle of luminiferous and electrical radiation, we must abandon the idea that it can be used as a means of distinguishing quasi-absolute from relative motion. But we cannot stop there; failure to detect motion through the ether must not only be admitted, it must also be explained.

What we may call the conservative explanation seeks to preserve the ether and the associated idea of quasi-absolute motion, and at the same time to get rid of the paradox that the observed velocity with which light-waves separate from their source is independent of whether that source is at rest or moving. It was offered first by FitzGerald of Dublin, and shortly afterwards (1895) by the great Dutch physicist H. A. Lorentz. It accounts for Michelson's failure to detect motion through the ether by the brilliantly simple supposition that the effects are masked by a contraction of the apparatus along the line of movement. In the experiment (it will be remembered) a light pulse emitted from O may be supposed to be reflected from two mirrors at A and B, OA and OB being of equal length but at right angles to one another. If the apparatus were at rest in the ether, the parts of the pulse reflected at A and B would return to O together; but if it were moving along the line OA with uniform speed v with respect to the ether, then the time taken for the double journey $OA + AO$ should be β times the time taken for the journey $OB + BO$, where $\beta = 1/\sqrt{(1 - v^2/c^2)}$. No such difference is observed; but it would obviously be cancelled out if the mere movement through the ether reduced OA to the length OA/β. We are to take it, then, that a contraction of this amount, affecting uniformly all

material bodies in motion through the ether, actually takes place.

Einstein's rival explanation (1905) looks infinitely less ingenious, for it consists merely in taking the Michelson and Morley result at its face-value. The experimenters could not find that the motion of the source made any difference to the measured velocity of the emitted light. Let us admit, then, says Einstein, that it actually makes no difference, and that the velocity of light is an "invariant" of nature, always the same from whatever standpoint it is measured. If that view ignores the rôle of the ether, so much the worse for the ether. The plain truth is that the ether, as hitherto conceived, blocks the path of progress. Once thought of as a quasi-rigid medium, with a calculable density and elasticity, it has recently been shorn of most of its mechanical properties. It must now lose the last—the assumed immobility in virtue of which a spurious kind of absoluteness has been conferred upon some motions of bodies to the detriment of others. Henceforward it must be recognized frankly that any motion imputed to a body is motion with regard to some system of reference which is deliberately taken *pro hac vice* to be at rest. The question as to whether a body is "really" in motion or at rest is a nonsensical one, and ought not to be asked. Motions with regard to different systems of reference will of course differ, but one is not to be thought of as more real than another. The only privilege one can have over another is to be capable of more simple description—as the motions of the planets are more simply described with reference to the sun than with reference to the earth.

And the ether? Well, as we proceed we shall find that

in Einstein's physics, as in Professor Alexander's metaphysics,* space and time cease to be purely inert entities and begin to take a hand, so to speak, in the world's work. The properties of space will then be practically those of the ether shorn of its mechanical properties. In brief, space will take on the function of the ether as the continuous medium in which physical interactions are transmitted. For instance, absolute rotation becomes more intelligible when the space in which it occurs is thought of in this way.†

* S. Alexander, *Space, Time and Deity* (2 vols., 1920: Macmillan).

† "Newton might . . . have called his absolute space 'ether'; what is essential is merely that besides observable objects, another thing, which is not perceptible, must be looked upon as real, to enable acceleration or rotation to be looked upon as something real". Einstein, *Sidelights on Relativity*, p. 17 (trs. Jeffery & Perrett, 1922: Methuen). Nothing more can be said in this book about the troublesome question of absolute rotation; it can only be suggested that the quotation just made contains the germ of a possible reply to Whitehead's criticism (*Principle of Relativity*, p. 87) that "the Einstein theory in explaining gravitation has made rotation an entire mystery".

CHAPTER II

THE RESTRICTED THEORY OF RELATIVITY

§ 3. EINSTEIN'S doctrine about absolute and relative motion is plain common sense, but its consequences, when it is taken seriously, are revolutionary and startling.

Let an observer S survey the world from a point O (fig. 1, p. 39) and refer its events, in so far as they are spatial, to three rectangular axes, OX, OY and OZ, the last named being perpendicular to the paper at O; and in so far as they are temporal, let him refer them to an impeccable clock at his side. In order that occurrences at a distance may be properly dated, let space be sown with an infinite number of clocks, all visible from O and keeping time with the clock there. The question how S can be sure that this last condition is fulfilled is an important one. It can be answered in two ways. If the clocks form a practically continuous series in all directions from O, the observer can engage a gigantic army of demons to explore the field and to assure him that the clocks that are (practically) at the same place are telling the same time. Since some of the clocks in any one "place" will naturally be counted as among the clocks also in the next place, the synchrony of the whole set can thus be guaranteed.

Einstein's own method follows a different principle but leads to the same result. Let a light-signal be sent from O, and after reflection at the face of a distant clock

return to the same point. We may suppose that the light, on reaching the distant clock, illuminated its dial for a moment and so indicated the time of arrival. Then if the time thus recorded is exactly half-way between the time at which the signal was sent out from O and the time at which it returned there, the two clocks will be taken to be synchronous.

Note that the clocks seen from O in a single momentary glance will not all *appear* to be keeping the same time. For instance, let clocks at distances from O of 3×10^5 km., 6×10^5 km. and 9×10^5 km. be visible together in S's telescope. Then since light travels at 3×10^5 km. per second, the times shown by these clocks, if they are really synchronous, must be respectively 1, 2 and 3 seconds behind the clock at O; and this must be the case in whatever direction the clocks lie. In other words, we assume the principle of constant light-speed without reference to the question whether the coordinate system with its attached clocks is moving or at rest.

Next let S' be a second observer stationed at a point O' which is moving along OX with uniform velocity v with regard to O.* Let S' refer the world's events to axes $O'X'$, $O'Y'$ and $O'Z'$, parallel to the corresponding axes of the S-system, and let him carry along with him a standard clock at O' (of precisely the same pattern as S's) and a multiplicity of clocks scattered through space, synchronous with his standard clock and rigidly connected with his coordinate axes. Further, let S' take the opportunity, at the moment when O' coincides with O,

* To S' it will, of course, appear that O is moving along the axis with velocity $-v$. The two statements are to be regarded as precisely equivalent.

(i) to see that his clock is in agreement with S's, and (ii) to take from S a measuring rod which is an exact duplicate of the one which that observer intends to use in measuring distances in his system.

Lastly, suppose that at the moment when O and O' coincide a spark of extremely short duration is emitted from the common origin of coordinates. A thin wave will expand into the surrounding space, and as it reaches each of the clocks in the S-system and the S'-system, will mark the time of its arrival by a momentary illumination of the face. At the instant represented in fig. 1 let it reach two clocks, one belonging to each system, which happen to be in contact at a point P directly above the point P' in the figure, and let S and S' note the time of arrival, each of them by his own clock at P. We will not beg an important question by assuming the two times to be the same, but will call them respectively t and t'—it being understood that when the spark was emitted the time by the clocks in both systems was zero.

Immediately after their common illumination the clocks at P will of course separate, since one is attached to the S-system, the other to the S'-system. But the observers will now have leisure to determine their positions, S by measuring the three distances OA $(= x)$, AP' $(= y)$, $P'P$ $(= z)$, and S' by measuring the corresponding distances $O'A$ $(= x')$, AP' $(= y')$, $P'P$ $(=z')$. (The points A, P' and P in the S'-system will have separated from the similarly lettered points in the S-system, but as the clocks marked P in the two systems are rigidly connected with the respective axes there will be no difficulty in identifying them.)

Now the two observers have, by hypothesis, been

spectators of two events *: one the emission of the spark at the moment when O and O' were coincident, the other the arrival of the light-wave at the coincident clocks at P. They have also noted the interval of time between those events—each observer by his own clock. Having further determined the coordinates of P, each in his own system, they are in a position to calculate, each for himself, the speed with which the light-wave travelled from the first event-particle to the second. In carrying out the calculations for them we must remember two things. (i) There is neither justification for saying nor meaning in saying that the S'-system has moved from the S-system rather than that the S-system has moved from the S'-system. Each observer has an equal right to the view that he has remained where he was and that the other observer has moved away from him. Thus, *even after the points O and O' have separated*, S will declare that the light started from O, S' that it started from O', and each of them will be equally well entitled to his assertion. (ii) By Einstein's fundamental principle, the velocity of the light as measured by the two observers will nevertheless have the same value c. We have then:

$$c = \frac{\sqrt{x'^2 + y'^2 + z'^2}}{t'} \quad \text{and} \quad c = \frac{\sqrt{x^2 + y^2 + z^2}}{t}$$

from which it follows that

$$c^2t'^2 - (x'^2 + y'^2 + z'^2) = c^2t^2 - (x^2 + y^2 + z^2) = 0 \qquad (3:1)\dagger$$

* Following Professor Whitehead we shall refer to such events as 'event-particles' in order to indicate that they occupy both an infinitely small space and an infinitely short time.

† The reader will observe that mathematical results throughout

Now, the argument which here involved the point P might have been applied to any point reached by the light-wave in its expansion; that is, it applies to all conceivable points in the spaces of the two observers. It is legitimate, therefore, to raise the question what general connexions exist between the values of the S'-coordinates and the values of the corresponding S-coordinates. It is obvious that they cannot have the same values; for at any moment the coordinate of O' along the axis OX is $x' = 0$ in the S'-system and $x = vt$ in the S-system. It follows from (3 : 1) that there must be a compensating difference between the values of at least one other pair of corresponding coordinates. Now, there appears to be no reason why the measurements of y' and y or of z' and z should differ from one another; accordingly the compensating difference sought must be found in the values of the time-coordinates. It will be proved in § 13 that this is the case, and that the only admissible correspondences between the coordinate measurements in the two systems are those given in the formulæ:

$$\left.\begin{aligned} x' &= \beta(x - vt) \\ y' &= y \\ z' &= z \\ t' &= \beta(t - vx/c^2) \\ \text{where} \qquad \beta &= 1/\sqrt{(1 - v^2/c^2)} \end{aligned}\right\} \qquad (3:2)$$

For the present it will suffice to verify that this set of substitutions (generally referred to as the "Lorentz

the book are referred to by means of two numbers, of which the first is the number of the article, the second the number of the result obtained in the course of that article. For example (24 : 5) means the fifth numbered equation or formula in § 24.

transformation") actually brings the two expressions in (3 : 1) into agreement. Omitting the equal coordinates, we have:

$$c^2t'^2 - x'^2 = c^2\beta^2\left(t - \frac{vx}{c^2}\right)^2 - \beta^2(x - vt)^2$$
$$= c^2t^2\beta^2\left(1 - \frac{v^2}{c^2}\right) - x^2\beta^2\left(1 - \frac{v^2}{c^2}\right)$$
$$= c^2t^2 - x^2$$

§ 4. We described the results of this argument, by anticipation, as revolutionary and startling, and in truth they are. It is a familiar fact that the events of the world present different faces to observers in relative motion: for example, that they are not the same when viewed from a moving train as when seen from a signal-box window. But before Einstein no one ever supposed that they were not set in the same framework of spatial and temporal relations. The world, it was held, is spread out in a single space and moves down the ages in a single "corridor of time"; but we are now called upon to surrender that too simple notion. A given spatio-temporal arrangement of the world's events has validity, we learn, only for a particular group of reference-systems which are at rest with respect to one another, and for any reference-system in motion with regard to these the spatio-temporal arrangement is different. In other words, time flows in a different manner for observers in relative motion, and each observer sees the world in a spatial setting which corresponds with his special kind of time-flow.

Some consequences of this radical change of view will be investigated later. Meanwhile, it may be helpful to illustrate by a simple example the use of the Lorentz

transformation and how the results obtained by it differ from those previously accepted in mathematics.

Let a projectile be discharged horizontally with velocity u from the top of a tower. Then if y and x denote the distances it travels vertically downwards and horizontally in time t, we have:

$$y = \tfrac{1}{2}gt^2, \qquad x = ut; \qquad \text{whence} \qquad y/x^2 = g/2u^2.$$

Now suppose photographs to be taken of the trajectory by means (i) of a stationary camera and (ii) of a camera in a train moving with steady speed v parallel to the initial direction of the projectile, the plates being in both cases parallel to the plane of flight. The picture developed upon the first plate will be a reduced version of the parabola $y/x^2 = g/2u^2$; what will the other picture be?

According to the usual theory, the coordinates of the projectile as it appears from the train at time t are $y' = y = \frac{1}{2}gt^2$ and $x' = x - vt = (u - v)t$, so that the picture on the plate of the moving camera would be of the parabola

$$y'/x'^2 = g/2(u - v)^2 \qquad (4:1)$$

But according to the theory of relativity the proper substitutions to make are $y' = y = \frac{1}{2}gt^2$ and $x' = \beta(x - vt) = \beta(u - v)t$, leading to the parabola

$$y'/x'^2 = g/2\beta^2(u - v)^2 \qquad (4:2)$$

In all actual cases the velocity of the train is so small a fraction of the velocity of light that β is sensibly unity and the parabolas (4 : 1) and (4 : 2) would be indistinguishable. That is why the world has had to wait so long for the theory of relativity. But if trains had, like α-particles, acquired the habit of travelling at 2×10^4 km. per sec.,

it would probably have been worked out long ago. For in that case the value of y' corresponding to a given value of x' would be, according to (4 : 2), not, as for a train moving at 68 miles an hour, about 1 part in 10^{14}, but actually about 1 part in 225 greater than according to (4 : 1); and so large a discrepancy could hardly have remained unobserved and unexplained.

§ 5. In the case of the trivial problem just dealt with, it is clear that although for all practical purposes (4 : 1) and (4 : 2) are equivalent, only the latter gives a solution theoretically correct; for it is the only formula which holds good for all values of v. That remark leads to an extremely important generalization—one which Einstein dignified in his first classical memoir (1905) with the name the "Principle of Relativity". Expressed negatively, it asserts that if a mathematical law which claims to describe the behaviour of a physical system does not hold good for any two systems in uniform motion with regard to one another, it cannot be true; put into positive terms, it states that any such law must preserve its mathematical form when transformed in accordance with the substitutions in (3 : 2) from the system in which it is first formulated into any other system moving uniformly with regard thereto.

In Chapter IV we shall examine instances of the application of this fundamental principle. It will there be shown, for example, that the principle of the conservation of momentum expressed by the ordinary formula

$$\Sigma\, mu = \text{const.}$$

cannot be precisely true; for if we transform it from a system in which it is assumed to hold good to another

system in uniform motion with regard thereto, it changes its form. But our investigation will not stop at that unsatisfying conclusion. It will further be shown that if the mass of a body is not regarded as constant but as varying in accordance with the formula

$$M = m/\sqrt{(1 - u^2/c^2)}$$

where m is an absolute constant and u the velocity of the body in the reference-system, the law becomes universally true in the form ΣMw = const., where w is the velocity measured in any prescribed direction. Thus the application of the principle of relativity brings to light a fact which the older physics had never suspected. It is, of course, in a line with and doubtless connected with Sir J. J. Thomson's discovery that the "apparent" mass of an electrically charged particle depends in part upon its velocity.

In a later chapter (Ch. VII) a still more striking instance will come before us. We shall find that Newton's law of gravitation, like the ordinary principle of the conservation of momentum, cannot be exactly true because it does not survive transformation from one coordinate-system to any other. And in this case also the search for a correction which will make the law universally valid led Einstein to discoveries of fact of the utmost interest—including the famous discovery of the bending of light near the sun.

A third instance, although of less general interest, is worth citing if only because it was Einstein's starting-point in the epoch-making memoir of 1905. Modern electro-magnetic theory, as the reader knows, is based upon the differential equations of Clerk Maxwell in the

modified form given to them by Hertz. Now, it can be deduced from these equations that if a magnet is moved in the neighbourhood of a conducting circuit, an electric field will be created around the magnet and will set up a current in the conductor; but if the magnet remains still and the conductor is moved, though indeed the current will appear as before, no electric field will be produced around the magnet. It is, however, difficult to believe that Nature would actually behave in this one-sided way, distinguishing arbitrarily between the motion of the magnet and the motion of the conductor; it becomes, therefore, a matter of much theoretical interest to determine how the lack of symmetry arises. Examination shows that it springs from the idea, which we have now definitely discarded, that motions with regard to the ether are on a different footing from other motions. The Maxwell-Hertz equations involve not only space and time measures and the components of electric and magnetic forces, but also c, the velocity of electro-magnetic (including luminiferous) radiation; it was assumed, therefore, that they held good only for a system at rest in the ether and must take a different form in the case of a system in motion. But according to the principle of relativity, if they are true for any one system they must be true, in the original form, for any system moving with uniform velocity with regard to the former—substitutions for the space and time measures having been duly made in accordance with the Lorentz transformation. When Einstein applied this principle he found that the electric and magnetic forces grouped themselves in the equations in such a way that the discrepancy we have referred to disappeared. This happy result must be regarded as

strongly confirmatory of the soundness of his whole argument.

§ 6. The reader will see from the foregoing examples how wide of the mark is the common idea that Einstein has shaken the once firm foundations of physical science by proving that we have only "relative" where we fondly thought we had "absolute" truth. Einstein has, it is true, shown that the old view of the world was too simple: that its events are not contained in "two great common receptacles" of space and time, but exist in an endless variety of modes of spatio-temporal connexion. But a spatio-temporal system is not unreal simply because it turns out that there is a multiplicity of them instead of only one; it might as well be argued that the number two cannot be real because thirty things can be counted in threes and in fives as well as in pairs. Nor has the admission of the multiplicity of spatio-temporal systems destroyed the unity of nature or the universality of physical law. On the contrary, it has suggested, as we have just seen, a criterion of physical truth more searching and effective than any we possessed before, by means of which men have already reached a fuller and more exact understanding of some of the fundamental aspects of nature.

§ 7. Another word must be added to explain the title of this chapter. What Einstein called in 1905 *the* principle of relativity is now called the "restricted" (or the "special") principle in reference to its limitation to coordinate-systems in *uniform* relative motion. In transcending this limitation and in applying the principle to systems having *any* kind of relative motion, we pass from the "restricted" to the "general" theory of relativity.

CHAPTER III

THE GENERAL THEORY OF RELATIVITY

§ 8. Imagine a wide, featureless plain, a rain-storm in which the drops fall vertically with uniform but not necessarily equal velocities, and an airship, now hovering, now moving horizontally. So long as the airship is at rest above the ground the tracks of the raindrops will appear to a passenger as vertical straight lines, but if it begins to move above the plain with a steady speed the lines will slope from the vertical at different angles in accordance with the velocities of the different drops. What will happen if the uniform speed of the ship is exchanged for a uniform acceleration? The obvious answer is that the drops will now sweep by in parabolas with horizontal axes, the wider curves being followed by the faster, the narrower by the slower drops.

This answer is based upon two familiar facts. The first is the natural tendency of an observer to regard himself as at rest and to impute any motion he sees to the bodies around him; the other is that the path of a particle moving with uniform velocity in one direction and uniformly accelerated in a perpendicular direction is necessarily a parabola. The second fact is of course exemplified whenever a body is thrown into the air from the earth's surface. It has a uniform horizontal velocity due to the impulse with which it was projected and a

constant downward acceleration which we attribute to the uniform " field of force " of the earth's attraction.

Now, if, owing to the featureless character of the plain and the smoothness of his passage, our passenger was unable to discover that he was moving, he might well believe that the horizontal acceleration as well as the vertical velocity actually belonged to the raindrops. In that case he would infer that he had strayed into a region where there was a horizontal " field of force " which imposed upon all bodies free to move a constant acceleration equal and opposite to his own actual but unperceived acceleration with reference to the ground.

Again, imagine an observer shot into the air inside a transparent ball and pursued by rockets and other projectiles. All of these (air-resistance being left out of account) will be subject to the vertical acceleration g which characterizes the earth's field of force. But since exactly the same acceleration affects the ball also, its existence will be entirely concealed from the observer, and the companion projectiles will seem to him to be moving not in parabolas but in variously sloping straight lines—upwards or downwards, and faster or slower, in accordance with the differences between their original speeds of projection and his own. If he had wholly lost the sense of his motion and took these appearances at their face-value, he would conclude that he had passed into a region beyond the earth's attraction—a region, that is, where there was no field of force.

With these fantastic instances in mind, let us apply unflinchingly the principle that motions may be referred with equal legitimacy to any coordinate-system, extending that principle to include not merely systems in

uniform relative motion but any systems whatever. Then we see that it is no longer possible to admit the objective existence of uniform fields of force, such as the one supposed to exist immediately above a limited part of the earth's surface. The sole criterion of the existence of such a field is the existence of a uniform acceleration; and we have seen that uniform acceleration may be created or destroyed by mere motion of the system of reference. But there is no means of judging whether motion of a reference-system is "real" or only "relative"; the question is, as we have seen, a senseless one, and *any* reference-system may legitimately be regarded as at rest. It follows that the apparent field of force created by a motion of the reference-system is as good and "real" as any other; in other words, that no uniform field of force is "real" at all. In fact, force, regarded as a potential pull lying in wait to seize upon a body and drag it through space, must be relegated with the old-fashioned ether to the limbo of mathematical fictions.*

§ 9. It is vital to note that the argument just exemplified can be applied only to a *uniform* field of force. Suppose thousands of guns to discharge millions of projectiles with velocities varying in direction and amount but great enough to allow them to escape from the earth and become independent denizens of the solar system; and let our much-tried observer accompany one of them. As he travels through space his acceleration (with reference to the sun) will constantly change, but since the projectiles near him will always be subject to the same acceleration as his own, they will appear to be, some at rest, others

* This idea is, of course, as old as the earliest works of Mach (1872) and Kirchhoff (1874).

passing this way or that with uniform speed along straight lines. But the acceleration with regard to the sun of more distant members of the swarm will be substantially different from his. These will possess, therefore, an outstanding acceleration with regard to his projectile, and their tracks will be seen as curves exhibiting what he would have called in his unregenerate days the action of a varying force.

It is clear from this argument that the gravitational field around the sun cannot be disposed of by the method described in the preceding article. The kind of field there considered is characterized by the fact that the tracks of particles moving freely through it either are all straight lines, traversed with uniform speed, or can be reduced to such by a suitable movement of the observer. Fields possessing this character are conveniently called "Galilean"—the reference being to Galileo's law that uniform rectilinear motion implies the absence of external force. In distinction from Galilean fields, the kind of field studied in the present article is called a "permanent" gravitational field. We have seen that a limited part of it, immediately surrounding the observer, may be regarded as Galilean—the assumption that this is possible is called the "Principle of Equivalence"—but that no motion of the observer will eliminate the accelerations throughout its whole extent. The amount and distribution of the accelerations will appear different to different observers according to their relative motion; but accelerations of some kind will always be there.

§ 10. If the old idea of a force of attraction is taboo, how are we to account for a permanent gravitational field? Einstein's reply is that we must regard the

irreducible accelerations as expressing intrinsic characters of space and time around the "attracting" body. According to the old conception, space bears no responsibility for anything that happens in it; the sun and the sun alone is accountable for the planets' behaviour. According to the new conception, space itself has quasi-physical properties correlated with those of the "matter" immersed in it. It is those properties, not the sun's "action at a distance," which determine the behaviour of bodies free to move in the permanent gravitational field. In other words, and as we hinted in § 2, the ideas of space and ether have come very close to one another.

Professor Eddington has hit off the spirit of the new conception by a delightful analogy which we will not spoil by repetition *; we may, however, attempt a less lively one. Think of a huge sphere of jelly, with a golden ball at its centre and so made that its consistency (and therefore its refractive index) increase from the circumference inwards according to some regular law; and let a Newtonian light-corpuscle enter it along the prolongation of one of its chords. According to the discarded theory of the great thinker, the corpuscle would constantly swerve towards the centre and would therefore pass through the sphere along a curve instead of along a straight line. An observer might well attribute its behaviour to an attraction by the golden ball, but we should know that its movement was determined from point to point by the intrinsic character of the jelly. To complete the analogy, we must suppose that the golden ball is not an adventitious ornament inserted by the cook, but that the varying consistency of the jelly is somehow an expression of its

* *Space, Time and Gravitation*, p. 95.

presence and its nature. In much the same way the gravitation of a wandering comet towards the sun is to be thought of as determined by the properties of space, though these properties cannot be dissociated from the presence of the sun and are an inevitable expression of its nature as a material body.

§ 11. The reader will now understand that the law of gravitation which Einstein offers as a substitute for Newton's is a law about the metrical properties of space around the " attracting" mass. Since it is to have universal validity, it must be a mathematical formula whose form is preserved when it is transformed from any one system of coordinates to any other; and since each system has its own time-measure as well as its own space-measures, time as well as space must be involved in the metrical properties with which the law deals.

§ 12. We have now carried the description of Einstein's main ideas as far as it is profitable to go without the aid of mathematics. But before beginning to fill in some details of the sketch we must refer briefly to a theory of relativity different in important respects from the one expounded in these pages. In a triad of very notable books * Professor A. N. Whitehead has analysed the fundamental notions of time, space and matter with unprecedented care and profundity, and, while making full use of the " magnificent stroke of genius " by which Einstein and Minkowski transformed the old conceptions of space and time, has found himself compelled to take up a critical attitude towards some of Einstein's methods

* *The Principles of Natural Knowledge* (1919), *The Concept of Nature* (1920), and *The Principle of Relativity* (1922): all published by the Cambridge University Press.

and conclusions. The most salient difference between them is that Whitehead refuses to follow Einstein in attributing physical properties, and therefore heterogeneity, to space. It is a cardinal article of his philosophic faith that temporal and spatial relations must be uniform in character, and that if we assume the contrary we surrender the basis which is essential for the knowledge of nature as a coherent system. But uniformity is not the same thing as uniqueness; there are endlessly numerous time-orders depending on differences in the circumstances of motion of the observer, and there is for each time-order a corresponding space. Logically, time-order is prior to space-order; for space-order is merely the reflection into the space of one time-system of the time-orders of alternative time-systems.* The older physics was right, then, in treating physical phenomena as "contingencies" superimposed upon the uniformity of time and space. Nevertheless, Einstein is right in contending that laws expressing their character and connexion cannot be true unless they preserve the same mathematical form in all time and space systems.

Einstein, as we shall see later, regards as tests of the validity of his law of gravitation the facts (i) that it agrees approximately with Newton's and (ii) that it predicts three "crucial phenomena" which cannot be deduced from the Newtonian hypothesis—of which two at least, including the famous eclipse phenomenon, have been found to exist on the predicted scale. As regards the second point, it is interesting to note that Professor Whitehead's theory leads to exactly the same predictions, so that experience

* *The Principle of Relativity*, p. 8. Alexander (*Space, Time and Deity*, i, pp. 50–8) has much the same idea.

has produced, so far, no criterion by which the claims of the rival theories may be decided. On the other hand, the younger theory points to the existence of minute phenomena which do not appear to be deducible from the older; it is possible, therefore, that observation may one day give its verdict in favour of one rather than the other.

Nothing more can be said in this book about Dr. Whitehead's views; for our purpose is to expound Einstein's, and these, from the point where they become more interesting, diverge so widely from Whitehead's that it would be merely confusing to attempt to keep both sets in mind together. There can, however, be no question that Whitehead, in his wonderfully acute and convincing analysis of the fundamental presuppositions of physics—a work that will ever redound to the credit of British thought—and in the theory of relativity he has based on it, has formulated a body of doctrine with which the orthodox relativists must somehow come to terms.

CHAPTER IV

THE LORENTZ TRANSFORMATION AND SOME APPLICATIONS

§ 13. It is not necessary to repeat the argument by which in § 3 we reached the equations

$$c^2t'^2 - (x'^2 + y'^2 + z'^2) = c^2t^2 - (x^2 + y^2 + z^2) = 0 \tag{13 : 1}$$

as an interpretation of the result of the Michelson–Morley experiment. It is, however, desirable to offer a proof that the Lorentz transformation (3 : 2) not only satisfies these equations but is the only set of substitutions that will do so. Readers whose consciences will allow them to be contented with the simple verification given in § 3 may pass on to the next article.

(i) The first thing is to show that the S'-coordinates are all *linear* functions of the S-coordinates and conversely. Add to the S'-system (fig. 1) other systems, S'', S''', S^{iv}, etc., all related to the S-system in the way described in § 3 but moving along the common x-axis with different velocities. Then by the fundamental Principle of Relativity (§ 5), one and the same set of formulæ must govern the transformation of the coordinate measurements from any one of these systems to any other. Consider, for example, the x-coordinate in the first three systems, assuming the velocity of the S''-system to be

u relative to the S'-system and U relative to the S-system. Then we have

$$x' = f(x, y, z, t, v) \tag{13:2}$$
$$x'' = f(x', y', z', t', u) \tag{13:3}$$
$$x'' = f(x, y, z, t, U) \tag{13:4}$$
$$x = f(x'', y'', z'', t'', -U) \tag{13:5}$$

and so on; f being the same function in each case. The

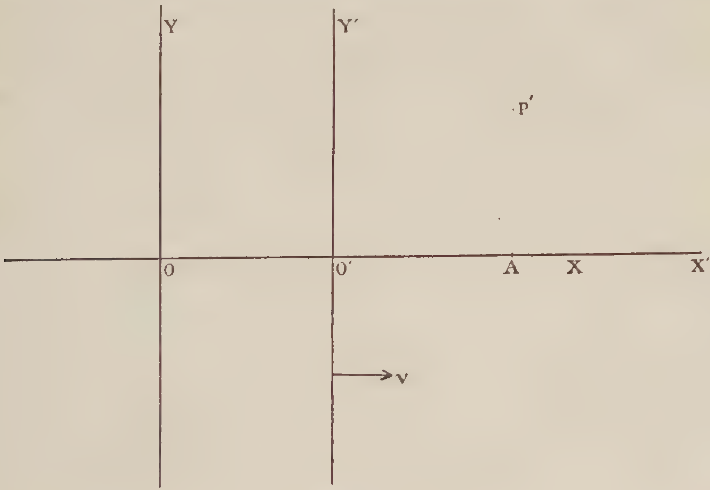

FIG. 1.

corresponding expressions for the other coordinates will, of course, involve other functions which may be symbolized as g, h, k.

If we substitute for x', y', etc., in (13 : 3) and equate with (13 : 4), we have

$$f\{f(x, y, \ldots), g(x, y, \ldots), h(x, y, \ldots), k(x, y, \ldots), u\} = f(x, y, z, t, U)$$

Now this relation is obviously possible if the several functions involve only the first powers of the variables, but in no other probable case. For instance, let the function f involve x^2 but no higher power of x. Then since the symbol f operates twice on the left-hand side it will produce there a term containing x^4, while on the right-hand side there is no such term. Consequently f must be a linear function. And the same argument applies to the others.*

(ii) Let us deal now with y'. It is evidently a function of y and v only. For whatever may be the values of x, z, t, its value is zero if the value of y is zero, and any connexion of the form

$$y' = y \, . \, F\,(x, z, t, v)$$

is ruled out by the condition that the function must be linear. Thus we can write

$$y' = \phi\,(v) \, . \, y \qquad (13:6)$$

where ϕ is a function of v to be determined. To determine it, note first that the magnitude of y' corresponding to a given magnitude of y cannot depend upon whether the S'-system moves forwards or backwards along the common axis. Hence

$$\phi\,(-v) = \phi\,(v)$$

Note again that the connexion remains unchanged if we

* Strictly, the argument as here given rules out only functions of the form $x' = a_0 + a_1x + a_2x^3 + \ldots + a_nx^n$; but when one considers that the number of transformations (i.e. the number of moving systems) may be increased indefinitely, and that the argument holds however many they be, there seems little room for doubt that its conclusion is true without qualification. The linearity of the transformation can also be proved from the uniformity of space, but the argument given in the text is more instructive.

conceive the S-system to be moving with velocity $-v$ instead of the S'-system's moving with velocity v. Hence

$$y = \phi(-v) \cdot y' = \phi(v) \cdot y' \qquad (13:7)$$

If (13 : 6) is now divided by (13 : 7) it appears at once that $y' = y$.

The same argument could evidently be used to prove that $z' = z$.

(iii) Next consider x' and t'. It is clear that x' cannot depend upon y or z; for if any point be taken on the $Y'Z'$-plane in fig. 1, we have for that point $x' = 0$ and $x = vt$, whatever may be the values of y and z. Thus x' can involve only x and t.

The same thing holds good for t'. For at time t' a point in the S'-system may have any values whatever for its y' and z'. But by (ii) these are equal respectively to y and z. Thus t' is independent of y and z.

We may therefore assume

$$x' = mx + nt \qquad \text{and} \qquad t' = px + qt \qquad (13:8)$$

where m, n, p, q are independent of the coordinates. But at O' in fig. 1 we have $x' = 0$ when $x = vt$. Hence the first equation becomes

$$x' = m(x - vt)$$

Substituting for x' and t' in (13 : 1), and remembering that $y' = y$ and $z' = z$, we obtain

$$c^2 (px + qt)^2 - m^2 (x - vt)^2 = c^2t^2 - x^2$$

Equating coefficients of x^2, t^2 and xt gives:

I. $p^2c^2 - m^2 = -1$; II. $q^2c^2 - m^2v^2 = c^2$;

III. $pqc^2 + m^2v = 0$.

From III it is clear that if v is positive p and q are of opposite signs. From I and II we get

$$p^2 = (m^2 - 1)/c^2 \text{ and } q^2 = (m^2v^2 + c^2)/c^2$$

while substitution in III yields

$$m^2 = q^2 = c^2/(c^2 - v^2) \text{ and } p^2 = v^2/\{c^2(c^2 - v^2)\}$$

Now if in (13 : 8) we let $t = 0$ in the first formula and $x = 0$ in the second, it becomes clear that m and q are both positive. Hence, putting β for $1/\sqrt{(1 - v^2/c^2)}$ we have

$$m = q = \beta \text{ and } p = -\beta\, v/c^2$$

so that the formulæ of (13 : 8) become

$$x' = \beta\,(x - vt) \text{ and } t' = \beta\,(t - vx/c^2)$$

Gathering the results together we have, for the Lorentz transformation,

$$\left.\begin{aligned} x' &= \beta\,(x - vt) & x &= \beta\,(x' + vt') \\ y' &= y & y &= y' \\ z' &= z & z &= z' \\ t' &= \beta\,(t - vx/c^2) & t &= \beta\,(t' + vx'/c^2) \end{aligned}\right\} \quad (13:9)$$

The right-hand values follow from those on the left in accordance with the principle of relativity, $-v$ being substituted for v.

§ 14. *Relativity of Length and Time.*—(i) Let $O'A$ in fig. 1 represent a rod at rest in the S'-system and therefore moving with speed v in the S-system. Let its length be l' in the S'-system and l in the S-system. Then, since $l' = x'$ and $l = x - vt$, we have by (13 : 9)

$$l' = \beta l \qquad \text{and} \qquad l = \alpha l'$$

where

$$\alpha = 1/\beta = \sqrt{(1 - v^2/c^2)} \qquad (14:1)$$

Thus the moving rod, viewed from a point stationary in the S-system, would appear to be shortened, the amount of apparent contraction depending on the speed. When $v = c$, $a = 0$; from which we deduce that a body flying with the speed of light away from an observer would seem to him to have no thickness at all.

(ii) Put $x' = 0$ in the equivalence $t = \beta(t' + vx'/c^2)$. Then it appears that the interval since O and O' were coincident, which is measured as t' by the clock at O', is measured as $\beta t'$ by the clock at O. In other words, the rate of the clock at O' is only a times the rate of the clock at O. Hence, as Langevin pointed out, if an observer were shot from the earth with a speed only a little inferior to that of light and returned to it after (say) a century of terrestrial time, the total time of the journey as measured by his own clock might be only a day or two. If he travelled with the full speed of light, time would stand still for him, for in that case a would be zero.

About these paradoxes some more will be said later (p. 57). Meanwhile it should be observed that the apparent contraction of the rod $O'A$ as viewed in the S-system is not at all the same as the hypothetical FitzGerald–Lorentz contraction, though it has the same numerical measure. The FitzGerald contraction was a contraction affecting a rod in a system in which it was at rest, if that system happened to be in motion in the ether. Here the rod $O'A$ presents no contraction to the observer S', however fast he may be moving with regard to S or any other observer. What happens, according to the theory of relativity, is that a rod whose length is l' when it is at rest in a system, has

length al' when it moves longitudinally in that system with uniform velocity v.

§ 15. *Proper Time.*—Let a particle P move with uniform velocity u in any given space-system S. Take the x-axis of S parallel to the direction of u and regard P as the origin of axes of coordinates parallel to those of S. Let P be accompanied by a standard clock of the same construction as the one at the S-origin. Then if δT be any time-element measured at P and δt the corresponding element measured by the S-clock, we have by the preceding argument

$$\delta T = \delta t \sqrt{(1 - u^2/c^2)} \qquad (15:1)$$

If the velocity varies in amount or direction or both, let the x-axis of S be adjusted for each time-element. Then we have

$$T - T_0 = \int_{T_0}^{T} dT = \int_{t_0}^{t} \sqrt{(1 - u^2/c^2)}\, dt \qquad (15:2)$$

The integral $T - T_0$, measured from some epoch T_0, is called the "proper time" of the particle. Note that the proper time between two events in the particle's history is always less than the time between the same two events as measured by an observer who watches its behaviour from a standpoint with regard to which the particle is in motion.

§ 16. *Relativity of Velocity.*—In § 13 we pictured two systems S' and S'' both moving along the x-axis of the system S, the velocity of S'' being assumed to be u relative to S' and U relative to S. According to ordinary ideas, since the velocity of S' relative to S is v, we should have $U = u + v$. We are now to see that this is not the case.

Put $\beta' = 1/\sqrt{(1 - u^2/c^2)}$, and $\beta'' = 1/\sqrt{(1 - U^2/c^2)}$;

then by (13 : 9) and the principle of relativity, if x'' is the x-coordinate in S''

$$x'' = \beta'(x' - ut') = \beta'\beta\{(x - vt) - u(t - vx/c^2)\}$$
$$= \beta'\beta\{(1 + uv/c^2)\,x - (u + v)\,t\}$$

Also
$$x'' = \beta''(x - Ut)$$

Equating coefficients, we have

$$\beta'' U = \beta'\beta(u + v) \text{ and } \beta'' = \beta'\beta(1 + uv/c^2)$$

and the division of the first of these equations by the second gives

$$U = \frac{u + v}{1 + \dfrac{uv}{c^2}} \qquad (16:1)$$

Let P be any point moving in the S'-system with velocity u parallel to the x-axis. Then since P may be thought of as carried along in a system S'' which itself moves along the x-axis with velocity u relative to the S'-system, its velocity (U) in the S-system is given by (16 : 1).

The following is a briefer but less interesting proof of the same important relation. From (13 : 9) we have

$$\delta x = \beta(\delta x' + v\delta t') \qquad \delta t = \beta(\delta t' + \frac{v}{c^2}\delta x')$$

whence
$$\frac{\delta x}{\delta t} = \frac{\delta x' + v\delta t'}{\delta t' + \dfrac{v}{c^2}\delta x'} = \frac{\dfrac{\delta x'}{\delta t'} + v}{1 + \dfrac{v}{c^2}\dfrac{\delta x'}{\delta t'}}$$

But $Lt\,.\,\delta x/\delta t = U$, and $Lt\,.\,\delta x'/\delta t' = u$. Thus the above equation becomes

$$U = \frac{u + v}{1 + \dfrac{uv}{c^2}}$$

In the foregoing argument u and U are both parallel to the x-axis; let us call them, therefore, longitudinal velocities and rewrite (16 : 1) in the form

$$U_l = \frac{u_l + v}{1 + \frac{u_l v}{c^2}} \qquad (16:2)$$

Next let P have in the S'-system a transverse velocity u_t parallel to the y-axis, and let the corresponding velocity in the S-system be U_t. Then since $\delta y' = \delta y$ we have

$$U_t = Lt \frac{\delta y}{\delta t} = Lt \frac{\delta y'}{\beta\left(\delta t' + \frac{v}{c^2}\delta x'\right)} = Lt \frac{\frac{\delta y'}{\delta t'}}{\beta\left(1 + \frac{v}{c^2}\frac{\delta x'}{\delta t'}\right)}$$

$$= \frac{a u_t}{1 + \frac{u_l v}{c^2}} \qquad (16:3)$$

If z and z' be substituted for y and y' in (16 : 3), it is seen that if in the S'-system P has no velocity parallel to the z-axis, then it has also no velocity parallel to that axis in the S-system.

Lastly, let the velocity of P in the S'-system be u in a direction making an angle θ with the x-axis, and let its velocity be U in the S-system making an angle ϕ with the x-axis. Then in the above we can substitute $u_l = u\cos\theta$, $u_t = u \sin\theta$, $U_l = U\cos\phi$, $U_t = U\sin\phi$, and

$$U^2 = U_l^2 + U_t^2 = \frac{(u\cos\theta + v)^2 + \left(1 - \frac{v^2}{c^2}\right)u^2\sin^2\theta}{\left(1 + \frac{uv}{c^2}\cos\theta\right)^2}$$

whence

$$U = \frac{\sqrt{\{(u^2 + 2uv\cos\theta + v^2) - (uv\sin\theta/c)^2\}}}{1 + (uv/c^2)\cos\theta} \qquad (16:4)$$

It is particularly interesting to use (16 : 2) to find the velocity in the S-system of a pulse of light issuing from a point P carried along with velocity v in the S'-system; for in accordance with the principle of constant light-velocity the result should be c. In fact, putting c for u_l we have at once

$$U_l = \frac{c+v}{1+\frac{v}{c}}$$
$$= c$$

as ought to be the case. To bring out the point in a still more striking way, let the S'-system itself advance along the x-axis with the velocity of light. Putting $u_l = v = c$, we now have

$$U_l = \frac{c+c}{1+1}$$
$$= c$$

—a result flagrantly contradictory to "common sense" but entirely in agreement with the fundamental principle of the theory of relativity.

Finally we can prove that the sum of two velocities ever so little less than c is itself always less than c. For that purpose put $u_l = c - p$, $v = c - q$, where p and q are both positive. Then

$$U_l = \frac{2c-(p+q)}{1+\dfrac{c^2-(p+q)c+pq}{c^2}}$$
$$= c^2 \cdot \frac{2c-(p+q)}{2c^2-(p+q)c+pq}$$
$$= c \cdot \frac{2c^2-(p+q)c+0}{2c^2-(p+q)c+pq}$$
$$< c$$

These results lead to the idea that the velocity of light is not only unique in being the only velocity which is invariant for all systems, but is also a limiting velocity which the speed of material particles may approach but cannot exceed or even reach.

§ 17. *The Relativity of Mass.*—Among the solidest foundations of the old physics were the principles that the mass of a body is an unchanging quantity and that the sum of the momenta of a number of bodies in dynamic relations is constant. It is, however, distressingly easy to prove that these two generalizations, taken together, cannot be true.

Consider a number of masses in the S'-system in movement under one another's action, and let $\Sigma m = K_1$ and $\Sigma mu = K_2$, where m is the mass of a body, u its velocity, assumed to be parallel to the x-axis in the S'-system, and K_1, K_2 constants. Now, according to the older ideas about relative velocity, the velocity of a mass in the S-system would be $U = u + v$, so that we should have

$$\Sigma mU = \Sigma mu + \Sigma mv = \Sigma mu + v\Sigma m = K_2 + vK_1 = \text{const.}$$

Thus, if momentum were conserved in any one system it would also be conserved in any other system moving with uniform relative velocity. But by (16 : 1)

$$\Sigma mU = \Sigma m \frac{u + v}{1 + \frac{uv}{c^2}} = \Sigma \frac{mu}{1 + \frac{uv}{c^2}} + v\Sigma \frac{m}{1 + \frac{uv^2}{c^2}}$$

and the constancy of the momentum-sum disappears.

We are confronted, therefore, by two alternatives: either we must abandon one or both of the principles of conservation or else we must seek some way of expressing

them which will survive transformation from one system to another. Let us try the less desperate policy first.

A clue to a plan is offered by the suggestion that the behaviour of bodies in movement within a system is to be correlated with the flow of their own time, not the time of the system. Now by (15 : 1) the time of a body moving in a system with velocity u flows $\sqrt{(1 - u^2/c^2)}$ times as fast as the time of the system. Let us, then, try the effect of assuming that the real constants of nature are

$$\Sigma \frac{m}{\sqrt{\left(1 - \frac{u^2}{c^2}\right)}} = K_1 \text{ and } \Sigma \frac{mu}{\sqrt{\left(1 - \frac{u^2}{c^2}\right)}} = K_2 \qquad (17:1)$$

This involves the supposition that the mass of a body is not a constant but varies according to its velocity in a system, being m in a system where it is at rest and $m/\sqrt{(1 - u^2/c^2)}$ in a system where its velocity is u.

Now, in the S-system the formulæ for mass-sum and momentum-sum corresponding to those of (17 : 1) are

$$\Sigma \frac{m}{\sqrt{\left(1 - \frac{U^2}{c^2}\right)}} \quad \text{and} \quad \Sigma \frac{mU}{\sqrt{\left(1 - \frac{U^2}{c^2}\right)}}$$

and the question is whether the values of these expressions are constant if the values of those in (17 : 1) are constant.

We have
$$U = \frac{u + v}{1 + \frac{uv}{c^2}} = c^2 \frac{u + v}{c^2 + uv}$$

whence by easy algebra

$$1 - \frac{U^2}{c^2} = \frac{(c^2 - u^2)\ (c^2 - v^2)}{(c^2 + uv)^2} = \frac{\left(1 - \frac{u^2}{c^2}\right)\left(1 - \frac{v^2}{c^2}\right)}{\left(1 + \frac{uv}{c^2}\right)^2}$$

Thus

$$\Sigma \frac{m}{\sqrt{\left(1 - \frac{U^2}{c^2}\right)}} = \Sigma \frac{m\left(1 + \frac{uv}{c^2}\right)}{\sqrt{\left(1 - \frac{u^2}{c^2}\right)\left(1 - \frac{v^2}{c^2}\right)}}$$

$$= \beta\Sigma \frac{m}{\sqrt{\left(1 - \frac{u^2}{c^2}\right)}} + \frac{v\beta}{c^2} \cdot \Sigma \frac{mu}{\sqrt{\left(1 - \frac{u^2}{c^2}\right)}}$$

$$= \beta K_1 + \frac{v\beta}{c^2} \cdot K_2 \qquad (17:2)$$

Again

$$\Sigma \frac{mU}{\sqrt{\left(1 - \frac{U^2}{c^2}\right)}} = \Sigma \frac{m\,(u + v)}{\sqrt{\left(1 - \frac{u^2}{c^2}\right)\left(1 - \frac{v^2}{c^2}\right)}}$$

$$= \beta\Sigma \frac{mu}{\sqrt{\left(1 - \frac{u^2}{c^2}\right)}} + v\beta\Sigma \frac{m}{\sqrt{\left(1 - \frac{u^2}{c^2}\right)}}$$

$$= \beta K_2 + v\beta K_1 \qquad (17:3)$$

But the expressions reached in (17 : 2) and in (17 : 3) are both constant in value. Thus it has been shown that if mass-sum and momentum-sum are estimated in accordance with the formulæ of (17 : 1) they are constant in all systems in uniform relative motion.

For simplicity, the foregoing investigation has been confined to the case in which all particles are moving parallel to the x-axis; its result can, however, be generalized without difficulty. We must distinguish between the longitudinal and the transverse momentum sums and, assuming their constancy for the S'-system, must demonstrate it separately for the corresponding directions in the S-system. The assumption as regards mass will still be that $\Sigma m/\sqrt{(1-u^2/c^2)}$ is constant in the S'-system,

but u will now be the whole velocity of a particle. The first step is to show, using (16 : 2, 3), that

$$\begin{aligned} 1 - U^2/c^2 &= 1 - (U_l^2 + U_t^2)/c^2 \\ &= a^2\,(1 - u^2/c^2)/(1 + u_l v/c^2)^2 \qquad (17:4) \end{aligned}$$

This result may be used to prove, by the former method, that the longitudinal momentum is given by (17 : 3), with the modification that U_l and u_l must be substituted for U and u in the numerators of the fractions. For the transverse momentum we have by (16 : 3) and (17 : 4),

$$\Sigma \frac{mU_t}{\sqrt{(1 - U^2/c^2)}} = \Sigma \frac{mu_t}{\sqrt{(1 - u^2/c^2)}}$$

i.e. the formula is unaffected by transformation. Since both the longitudinal and the transverse momentum-sums are constant, it follows that if w is the velocity in any prescribed direction, then the sum $\Sigma mw/(1 - u^2/c^2)$ transforms into $\Sigma mW/(1 - U^2/c^2)$ on passage from the S'-system to the S-system and that both sums are constant.

To sum up. Every particle has a "proper mass" m, which is, so to speak, its mass in its own system, and is an invariable factor determining its behaviour in all its dynamical transactions. From the standpoint of a system in which the particle has a velocity u, the mass is a quantity M, connected with m by the relation

$$M = m\left(1 - \frac{u^2}{c^2}\right)^{-\frac{1}{2}} \qquad (17:5)$$

The principle of the conservation of mass takes the form that for systems in uniform relative motion ΣM is constant within each system, though varying from one system to another. Similarly the principle of the conservation of momentum is to be understood in the

sense that ΣMw is constant within each system, where w is the velocity of the particle in any prescribed direction.

§ 18. *Kinetic Energy*.—If the reader applies the method of § 17 to the expression for the kinetic energy of a particle, he will easily see that neither the formula $\frac{1}{2}mu^2$ nor the formula $\frac{1}{2}Mu^2$ survives transformation. We are therefore driven to conclude that neither of them can accurately express the energy a particle possesses in virtue of its motion. Now if we expand (17 : 5) we obtain

$$M = m + \tfrac{1}{2}(m/c^2)\,u^2 + \tfrac{3}{8}(m/c^4)\,u^4 + \ldots \qquad (18:1)$$

If the units of length and time are so chosen that the velocity of light is unity—a device frequently useful in the theory of relativity—the formula for M becomes

$$M = m + \tfrac{1}{2}mu^2 + \tfrac{3}{8}mu^4 + \ldots\ldots \qquad (18:2)$$

in which the terms, as is seen from (18 : 1), rapidly decrease in value unless u is nearly as great as the velocity of light. We draw from this result two inferences: (i) the proper mass of a particle is a quantity of the same kind as its kinetic energy, and (ii) the kinetic energy is only approximately expressed by the usual formula $\frac{1}{2}mu^2$.

It has been customary to regard a particle's kinetic as only part of its whole energy, the rest being thought of as "internal" energy. Formula (18 : 2) suggests that the internal energy is identical with the proper mass, and that the mass M which is conserved from system to system is the sum of the internal energy and the energy due to motion in the particular system. The principle of the conservation of mass thus becomes a special case of the principle of the conservation of energy.

CHAPTER V

THE SPACE-TIME INVARIANT

§ 19. Let an event-particle (e.g. the emission of a momentary spark or the production of a momentary noise) have x_1, y_1, z_1, t_1 for its coordinates in the S-system, and let another event-particle occur at x_2, y_2, z_2, t_2 in the same system. Let r be the distance between the points where the events take place. Then

$$r^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 \qquad (19:1)$$

Now by the Lorentz transformation (13 : 9)

$$\left.\begin{aligned} x'_2 - x'_1 &= \beta\,(x_2 - x_1) - \beta v\,(t_2 - t_1) \\ y'_2 - y'_1 &= y_2 - y_1 \\ z'_2 - z'_1 &= z_2 - z_1 \\ t'_2 - t'_1 &= \beta\,(t_2 - t_1) - \beta\frac{v}{c^2}\,(x_2 - x_1) \end{aligned}\right\} \qquad (19:2)$$

and the substitution of these values in (19 : 1) gives for the distance between the same two places in the S'-system

$$\begin{aligned} r'^2 = \beta^2\,(x_2 - x_1)^2 - 2\beta^2 v\,(x_2 - x_1)\,(t_2 - t_1) + \beta^2 v^2\,(t_2 - t_1)^2 \\ + (y_2 - y_1)^2 + (z_2 - z_1)^2 \qquad (19:3) \end{aligned}$$

Thus the distance between two points is not a constant, but varies from system to system.

But now subtract from r^2 the number $c^2\,(t_2 - t_1)^2$ and from r'^2 the equivalent for $c^2\,(t'_2 - t'_1)^2$ taken

from (19 : 2). Then the expression for $r'^2 - c^2(t'_2 - t'_1)^2$ becomes

$$\beta^2\left(1 - \frac{v^2}{c^2}\right)(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 - \beta^2(c^2 - v^2)(t_2 - t_1)^2$$

that is,

$$(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 - c^2(t_2 - t_1)^2 \quad (19:4)$$

since $\beta^2 = 1/(1 - v^2/c^2)$. Thus it appears that although the values of r^2 and r'^2 differ, the values of $r^2 - c^2(t_2 - t_1)^2$ and $r'^2 - c^2(t'_2 - t'_1)^2$ are the same. This result, which is the foundation upon which all our future work will be based, is expressed by saying that the "interval" or the "separation" between event-particles is an invariant for all systems in uniform relative motion. The word "interval" is the one usually employed, but has the disadvantage of putting unfair emphasis on the time-element in the quantity; we shall, therefore, use Professor Whitehead's term "separation," i.e. separation in both space and time.

Observe that $c(t_2 - t_1)$ is the distance that light would travel in the time-interval between the two events. In ordinary cases this is greater than the distance r between the places where the events occur. It is usual, therefore, to reverse the signs in (19 : 4), and to express the separation in the form

$$s^2 = -(x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2 + c^2(t_2 - t_1)^2 \quad (19:5)$$

Since this relation holds good for all pairs of event-particles it will hold good when they occur very near one another in space and time. In that case (19 : 5) is naturally written

$$\delta s^2 = -\delta x^2 - \delta y^2 - \delta z^2 + c^2\delta t^2 \quad (19:6)$$

If $c\ (t_2 - t_1)$ is less than r, the square of the separation as measured by (19 : 5, 6) becomes negative; that is, s becomes an "imaginary" number. It is, however, better to suffer this minor inconvenience than to have different formulæ for the different cases. It will be understood that the "imaginary" value of s corresponds to a perfectly real situation.

In the foregoing argument we have assumed, as usual, that the systems in uniform relative motion have a common x-axis along which the relative motion takes place. It is extremely important, therefore, to show that the invariance of (19 : 5) and (19 : 6) is by no means limited to cases in which that condition of affairs obtains. Let the S'-system be allowed to continue its original motion with regard to the S-system, but let its axes, while remaining rectangular, be shifted to a new origin and take up new alignments. These changes can make no difference whatever to the observer's estimate of the distance r' between the points where the two event-particles occurred; nor will it affect the rate of his clocks. It follows that, for him, the expression $-r'^2 + c^2\ (t'_2 - t'_1)^2$ will retain its value although his axes of coordinates are no longer parallel to those of the S-system. In the same way, the S-observer may change the origin and orientation of his axes without in the least affecting the value of the expression $-r^2 + c^2\ (t_2 - t_1)^2$. And since the two estimates of the "separation" between the event-particles were equal before these changes took place, they remain equal afterwards. But the two axis-systems may now have *any* relative orientation, and their relative velocity may make *any* angles whatever with the axes. Thus it appears that the separation between two event-

particles is an invariant in exactly the same way as the distance between two given points at rest in a single system: that is to say, it does not depend at all upon the disposition of the coordinate-axes which the observer may use in measuring it.

The upshot of these remarks may be expressed in the statement that the invariance of (19 : 5, 6) plays in the theory of relativity the part which Pythagoras's theorem plays in ordinary static geometry.

§ 20. Let (19 : 5) be written in the simpler form

$$s^2 = -r^2 + c^2T^2 \qquad (20:1)$$

where r is the distance and T the time-interval between the two event-particles; then consideration of the possible cases leads to instructive results.

(i) Let s^2 be negative, and let the line joining the two event-particles be taken as the common x-axis, so that $r = (x_2 - x_1)$.

Then we have $$T^2 < \frac{r^2}{c^2} = \frac{u^2}{c^2} \cdot \frac{r^2}{c^2} \qquad (20:2)$$

where u is some number less than c, taken with the same sign as r so that ur may be positive.

Let $T = (t_2 - t_1)$ be positive, so that, by (20 : 2),

$$T = \frac{u}{c^2} r \quad \text{and} \quad r = \frac{c^2}{u} T \qquad (20:3)$$

and let T' be the interval between the same pair of event-particles in the S'-system. Then since by (19 : 2)

$$T' = \beta \left(T - \frac{v}{c^2} r\right)$$

it follows that

$$T' = \beta T \, (1 - v/u) \qquad (20:4)$$

Now if u and v have the same sign, the sign of T' will depend upon whether v is greater or less than u. But the sign of T' determines whether the event which, by hypothesis, happens first in the S-system, happens first or second in the S'-system. Thus it appears that an event which occurs after another in the S-system may occur either after it or before it in the S'-system according to whether the relative velocity of the two systems is less or greater than u. If $v = u$ the events which, by hypothesis, are not simultaneous in the S-system, are nevertheless simultaneous in the S'-system. Since s is constant for all systems, it follows from (20 : 1) that the distance between two given event-particles is least in the system in which they occur simultaneously. In other systems the increase in their distance is compensated by an increase of the time-interval between them. Thus it appears that space may, in a certain sense, be regarded as convertible negatively into time, and conversely.

Since in the case now under consideration the time-order of two events may be different in different systems we cannot think of the events as causally connected with one another. This is part of what Professor Whitehead has in view when he defines event-particles for which $c^2T^2 < r^2$ as "co-present". The term also implies that if this relation holds between two event-particles, there is always a system in which they happen simultaneously—namely, the system for which $u = u$.

The fact that time and space are, as we have said, convertible, throws light upon the paradoxes of § 14. By the length of the rod $O'A$ is clearly meant the distance between two points which are occupied by its ends simultaneously. Let two momentary sparks be emitted

simultaneously at O' and A in the S'-system; then by the preceding argument they will not be simultaneous in the S-system. In fact, if r and r' are the distances between the two sparks in the S-system and the S'-system respectively, there is between them in the S-system an interval T such that

$$-r^2 + c^2T^2 = -r'^2 = s^2$$

since the separation between the two event-particles is the same in both systems. It follows that the observer in S, measuring the length of the rod by the distance between two points which its ends occupy simultaneously, cannot take the same two points as the observer in S', and therefore cannot make the length of the rod the same. Thus the paradox of § 14 (i) arises from the fact that S' has all of the separation between the two sparks in terms of length; while S has it partly in length and partly in time—or, rather, has more length because he has also some time!

Mutatis mutandis the same explanation applies to the apparent dilatation of time. Strictly speaking, we mean by a time-interval a difference between the readings of a clock at a single place. Now let the S'-clock signalize the beginning and end of one unit of time by emitting momentary sparks. These will occur in the same place in the S'-system but at different places in the S-system. The constancy of the separation is now expressed by the relation

$$-r^2 + c^2T^2 = c^2 = s^2$$

which shows that S will have more time than S' because the separation contains for him some length as well as time; and the paradox of § 14 (ii) arises herefrom.

(ii) Next let $c^2T^2 > r^2$; then u in (20 : 2) must be greater than c and therefore greater than any possible v. It follows that in (20 : 4) the factor $(1 - v/u)$ is always positive whatever the signs of u and v. Hence the sign of T' always agrees with the sign of T: that is, events which occur in a given order in one system occur in the same order in any other system. But since s^2 in (20 : 1) is positive, it is not possible for T to be zero in *any* system; in other words, simultaneity of events is absolutely excluded. It is, however, possible for two events to happen at the same place at different times. In Professor Whitehead's nomenclature, the events all belong to one another's "kinematic past" or "kinematic future".

(iii) Lastly, let $c^2T^2 = r^2$. Then it is clear in the first place that if either T or r is zero the other variable is also zero. That is, if two events happen at the same place in a given system they also happen at the same time in that system—and conversely. Members of a one-dimensional chain of event-particles among which the relation now considered obtains can never revisit a place once occupied or occupy two places at the same time.

In (20 : 2) and in (20 : 4) u is now the same (numerically) as c and thus necessarily greater than v. The signs of T' and T must, therefore, always be the same. It follows that the order of events is the same in all systems.

If A is the earlier and B the later of a pair of events whose distance is given, B may take place anywhere on a sphere whose radius is cT and therefore grows with the velocity of light. It will be a sphere for all systems since $c^2T^2 = r^2$ for all systems. For the same reason, the radius of the sphere will be the same at the same time in all systems. This is, in fact, the only one of the three cases

in which time and space are not convertible, and what happens in one system happens in exactly the same way in all others.

In Professor Whitehead's phraseology the event-particles considered in (iii) belong to one another's " causal past " or " causal future ". The terms imply the conception that causal activity spreads from a given point-event to others with the speed of light.

§ 21. *Space-Time*.—The mutual convertibility of space and time, to which we referred in the last article, makes it impossible any longer to think of them as absolutely distinct and separate features of the world. The first man clearly to realise this was H. Minkowski, who began a famous lecture by announcing that thenceforward space and time, considered in themselves, would sink to the position of mere shadows, and that only a kind of union of the two could claim independent existence. This " union " is the space-time whose properties we have just been examining. Whether, with some philosophers, we think of it as logically prior to events and in a sense generating them, or whether with others we regard it as an abstraction from events, does not for our present purpose matter. It is, however, essential to see that there is only *one* space-time, and that the indefinitely numerous systems of time and space which we recognize represent merely the different ways in which that one space-time may be divided up. Note, however, that so long as we keep in view the whole of it, we cannot divide it into space *and* time; every mode of division has a time-like aspect and a corresponding space-like aspect.

Provided we bear this truth in mind, it is of great help to conceive space-time in the light of analogies drawn

from our knowledge of its purely spatial aspect. Thus we may regard it as a continuum of four dimensions, of which three are space-like and the fourth time-like. In this four-dimensional continuum point-instants play the part which points play in three-dimensional space—or, if you prefer to put it so, event-particles replace the mere particles of the three-dimensional world. Correspondingly, the distance between points in space is replaced by the "separation" between point-instants or event-particles.

Conceived thus as a four-dimensional space, space-time ceases to contain anything of the nature of history—for past, future and present are all there together. For instance, the history of a particle which occupies different places in the world at different times becomes represented as a "world line" (the term is Minkowski's) of definite shape lying in the four-dimensional continuum just as a thin wire of definite shape may lie in three-dimensional space. If we mark two near point-instants upon a world line, their separation δs corresponds to the elementary distance δs between two near points on the curve in ordinary space. If we regard a particle as constituting the origin of its own system of reference, then for that system $x = y = z = 0$, and in (19 : 6) $\delta s = c\delta t$. The integral $\frac{1}{c}\int ds$ is thus what we have called (again following Minkowski) the particle's proper time (§ 15).

In accordance with the argument of § 20, we may take any point-instant P in the four-dimensional continuum and classify all the rest into (*a*) those which are co-present with it, (*b*) those which belong to its kinematic past or future, (*c*) those which belong to its causal past or future.

Since the last are characterized by the relation $s^2 = 0$ which holds between themselves and P, while the others are characterized respectively by the relations $s^2 < 0$ and $s^2 > 0$, the point-instants in (*c*) may be regarded as lying between the point-instants of (*a*) and (*b*). At any given instant the point-instants belonging to P's causal past and future lie, as we have seen, on the two-dimensional surface of a sphere; the whole aggregate of them may be regarded, therefore, as constituting a three-dimensional continuum of which one dimension is time-like. It is in this sense that Professor Whitehead speaks of P's causal past and future as a three-dimensional boundary between its co-present region and its kinematic past and future.*

§ 22. Some points in the analogy we are pursuing are brought out more clearly if in (19 : 6) we substitute for x, y and z the symbols u_1, u_2, u_2 and instead of $-ct$ write iu_4, where $i = \sqrt{-1}$. Equation (19 : 6) then becomes

$$-\delta s^2 = \delta u_1^2 + \delta u_2^2 + \delta u_3^2 + \delta u_4^2 \qquad (22:1)$$

which exhibits the separation as an (imaginary) quantity $i\delta s$ measuring the distance between two near points in a four-dimensional space. The total separation between any two event-particles A and B will be given by the integral

$$\int_A^B ids \qquad (22:2)$$

Take A as the origin of rectangular coordinates, and let B occupy a continuous series of positions on a world line. Then if, for every position of B, the value of the integral (22 : 2) is

$$\sqrt{(u_1^2 + u_2^2 + u_3^2 + u_4^2)} \qquad (22:3)$$

* *Theory of Relativity*, p. 30.

where the symbols are coordinates of B, then we shall have something analogous to what happens when, in space of two or three dimensions, the integral of δs is taken along a straight line. For in that case $\int_A^B ds$ is $\sqrt{(x^2 + y^2)}$ for two dimensions and $\sqrt{(x^2 + y^2 + z^2)}$ for three. We may speak, then, of a world line which fulfils this condition as being *straight*. Moreover, just as in the three-dimensional case a straight line is one along which all the differential coefficients, du_1/du_2, etc., have constant values, so here they will all be constant, including the differential coefficients with respect to the time-like variable u_4 and to s itself. It follows that if the value of the integral (22 : 2) is given by (22 : 3), the world line represents the *uniform* motion in a straight line which Galileo was the first to teach as characteristic of a particle free to move in a region devoid of force (§ 9).

There is, however, an important difference between a straight line in three-dimensional space and a uniform world line. If two points in space A and B are joined by any number of lines, the integral of δs, that is the length-integral, is least along the line which is straight; but if A and B are two point-instants in space-time, and we take the integral of δs, that is the separation-integral, along any number of world lines which include them, then the integral is *greatest* along the world line which is uniform in the sense just defined.

To see this, imagine the uniform world line to be broken up into an indefinitely large number of elements by point-instants in such a way that the separation δs between consecutive members of the series is the same all along

the line. Then the conception of uniformity implies that the distances δr between the consecutive points and the intervals δt between the consecutive instants of the series are also constant all along the line. That is, in the equation

$$\delta s^2 = -\delta r^2 + c^2\delta t^2 \qquad (22:4)$$

δs, δr and δt have the same values for each of the elements. Now consider a second world line which also includes the point-instants A and B, and let this be divided into elements by point-instants which are, as regards their time-aspects, simultaneous with the point-instants which mark the beginnings and ends of the corresponding elements of the uniform world line. Then in the equation for the separation between two members of the series of point-instants, δt is constant all along the second world line and equal to its value in the former series; but since the space-path between A and B is not straight, some at least of the distances between the points must be greater than the constant distances in the former series. If $\delta s'$ and $\delta r'$ refer to the second world line, it follows from the equation

$$\delta s'^2 = -\delta r'^2 + c^2\delta t^2 \qquad (22:5)$$

that $\delta s'$ must at least for some elements be less than the constant δs. For $c^2\delta t^2$ has the same value in (22 : 4) and (22 : 5), while $\delta r'^2$, we have said, is at least sometimes greater than δr^2. Hence the integral of the separation is a maximum along the uniform world line.

Observe from (13 : 1) that the separation integral along the uniform world line which exhibits the passage of light from A to B is always zero.

CHAPTER VI

SOME MATHEMATICAL NOTES

THIS chapter contains brief explanations of certain mathematical processes and results which will be used constantly in the sequel. The expert will not need them, but those whose mathematics are rusty may be glad to refresh their memories before plunging into the general theory.

§ 23. *Partial Differentiation.*—Differentiation in the theory of relativity is usually partial differentiation. Let the value of a variable v_1 depend upon the values of (say) four independent variables u_1, u_2, u_3, u_4. Let the value of u_1 change by a small amount δu_1 while the others remain constant, and let δv_1 be the resulting change produced in v_1. Then the limit of the fraction $\delta v_1/\delta u_1$ as δu_1 approaches zero is the partial differential coefficient of v_1 with respect to u_1, and is expressed by the symbol $\partial v_1/\partial u_1$. It measures the rate of change of v_1 per unit change in u_1 when the other variables are constant.

(i) Let u_a represent any one of the four independent variables, and let it suffer a small change δu_a; then the resulting change in v_1 is

$$\frac{\partial v_1}{\partial u_a} \delta u_a$$

that is: the rate of change of v_1 per unit change in u_a multiplied by the actual change in u_a which has taken

place. If all four independent variables suffer change, the total resulting change in v_1 will be the sum of the four independent changes:

$$\delta v_1 = \frac{\partial v_1}{\partial u_1}\delta u_1 + \frac{\partial v_1}{\partial u_2}\delta u_2 + \frac{\partial v_1}{\partial u_3}\delta u_3 + \frac{\partial v_1}{\partial u_4}\delta u_4$$
$$= \sum_a \frac{\partial v_1}{\partial u_a}\delta u_a \qquad [a = 1, 2, 3, 4] \qquad (23:1)$$

The meaning of the last line is that a is to receive in succession the values 1, 2, 3, 4, and that the four expressions thus obtained are to be added together.

As an example of the foregoing argument, let OU_1, OU_2 (fig. 2) be a pair of (not necessarily rectangular) axes, and

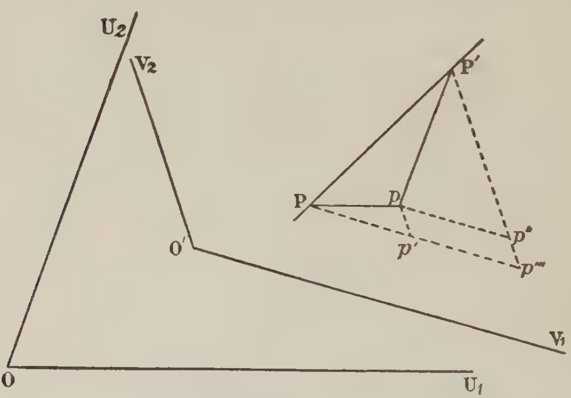

FIG. 2.

$$Pp = \delta u_1, \ pP' = \delta u_2;$$
$$Pp' = \frac{\partial v_1}{\partial u_1}\delta u_1, \ p'p = \frac{\partial v_2}{\partial u_1}\delta u_1; \ \ pp'' = \frac{\partial v_1}{\partial u_2}\delta u_2, \ p''P' = \frac{\partial v_2}{\partial u_2}\delta u_2;$$
$$\delta v_1 = Pp''' = Pp' + pp'' = \frac{\partial v_1}{\partial u_1}\delta u_1 + \frac{\partial v_1}{\partial u_2}\delta u_2;$$
$$\delta v_2 = p'''P' = p'p + p''P' = \frac{\partial v_2}{\partial u_1}\delta u_1 + \frac{\partial v_2}{\partial u} u_2$$

$O'V_1$, $O'V_2$ another pair in the same plane ; and let the coordinates of a point P be u_1, u_2 when referred to the first pair, v_1, v_2 when referred to the second.

Let P be moved to a neighbouring point P' by means of successive small displacements δu_1, δu_2, parallel respectively to OU_1 and OU_2. Each of these movements will, as the figure shows, cause a movement of P parallel to $O'V_1$ and also a movement parallel to $O'V_2$. Thus for the whole displacements parallel to these axes we shall have

$$\delta v_1 = \frac{\partial v_1}{\partial u_1}\,\delta u_1 + \frac{\partial v_1}{\partial u_2}\,\delta u_2$$

$$\delta v_2 = \frac{\partial v_2}{\partial u_1}\,\delta u_1 + \frac{\partial v_2}{\partial u_2}\,\delta u_2$$

Now imagine P to symbolize an event-particle whose coordinates in the two systems are respectively u_1, u_2, u_3, u_4, and v_1, v_2, v_3, v_4, while P' symbolizes another event-particle near the former in space and time. Then by an extension of the above results we have

$$\delta v_m = \frac{\partial v_m}{\partial u_1}\,\delta u_1 + \frac{\partial v_m}{\partial u_2}\,\delta u_2 + \frac{\partial v_m}{\partial u_3}\quad u_3 + \frac{\partial v_m}{\partial u_4}\,\delta u_4$$

$$= \sum_a \frac{\partial v_m}{\partial u_a}\,\delta u_a \qquad [m,\ a = 1, 2, 3, 4] \qquad (23:2)$$

where $[m,\ a = 1, 2, 3, 4]$ means (i) that there are four separate equations in which m is to have the values 1, 2, 3, 4 respectively, and (ii) that the right-hand side of each equation is a sum of four terms in which a has the values 1, 2, 3, 4 respectively.

(ii) Let fig. 3 represent a tract of country over which the barometric pressures at a given moment have been mapped with reference to the axes OU_1, OU_2; and let

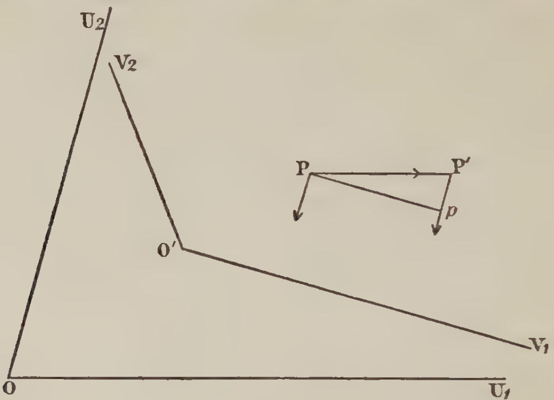

FIG. 3.

P be the pressure at the point P. From the map we could determine the pressure-gradient in any specified direction. For instance, the pressure-gradient parallel to OU_1 is the limit of $\delta P/\delta u_1$, where δu_1 signifies a small distance such as PP', and δP is the change of pressure along that distance. Thus the pressure-gradient in that direction is $\partial P/\partial u_1$.

Suppose we wish to calculate the gradient parallel to $O'V_1$, in terms of the gradients parallel to OU_1 and OU_2. To obtain it we can express the difference of pressure (δP) between the ends of a short distance Pp (Pp being parallel to OV_1) as the sum of the differences along the distances PP' and $P'p$. Thus we have

$$\delta P = \frac{\partial P}{\partial u_1}\,\delta u_1 + \frac{\partial P'}{\partial u_2}\,\delta u_2$$

where $\partial P'/\partial u_2$ is the pressure-gradient at P' parallel to OU_2. This equation, divided by δv_1, becomes

$$\frac{\delta P}{\delta v_1} = \frac{\partial P}{\partial u_1}\frac{\delta u_1}{\delta v_1} + \frac{\partial P'}{\partial u_2}\frac{\delta u_2}{\delta v_1}$$

If we now allow δv_1 to diminish, P' moves towards P, and we have as the limit

$$\frac{\partial P}{\partial v_1} = \frac{\partial P}{\partial u_1}\frac{\partial u_1}{\partial v_1} + \frac{\partial P}{\partial u_2}\frac{\partial u_2}{\partial v_1} \qquad (23:3)$$

Note that in specifying a barometric pressure we have to mention only its amount, but that in specifying a pressure-gradient we must also state its direction. A quantity of the first kind is called a "scalar", one of the second kind a "vector". A scalar quantity from which physical vector-quantities, such as velocity, electro-motive force, etc., can be calculated by partial differentiation is conveniently called a "potential". For instance, a scalar quantity is a "force-potential" if, given that its value is P at a specified point, we can infer that the force at that point in the direction u_a is $\partial P/\partial u_a$.

Let P be the value of a potential characterizing any point-instant whose co-ordinates are u_1, u_2, u_3, u_4 in one system and v_1, v_2, v_3, v_4 in another. Then by an extension of (23 : 3) we have

$$\begin{aligned}\frac{\partial P}{\partial v_m} &= \frac{\partial P}{\partial u_1}\frac{\partial u_1}{\partial v_m} + \frac{\partial P}{\partial u_2}\frac{\partial u_2}{\partial v_m} + \frac{\partial P}{\partial u_3}\frac{\partial u_3}{\partial v_m} + \frac{\partial P}{\partial u_4}\frac{\partial u_4}{\partial v_m} \\ &= \sum_a \frac{\partial P}{\partial u_a}\frac{\partial u_a}{\partial v_m} \qquad [m, a = 1, 2, 3, 4] \qquad (23:4)\end{aligned}$$

The reader will easily see that (23 : 4) is not valid only

for potentials, but holds good if P is *any* function expressible in terms of either set of coordinates.

(iii) Confine attention to the U-system in fig. 2. Let P and P' be the pressures at the two points P and P', let δs be the distance between the points, and let it be required to calculate the gradient in the direction PP' in terms of the gradients parallel to the axes. By the preceding argument we have

$$\delta P = \frac{\partial P}{\partial u_1}\,\delta u_1 + \frac{\partial P}{\partial u_2}\,\delta u_2$$

whence

$$\frac{\delta P}{\delta s} = \frac{\partial P}{\partial u_1}\frac{\delta u_1}{\delta s} + \frac{\partial P}{\partial u_2}\frac{\delta u_2}{\delta s} \qquad (23:5)$$

Now the displacement δs is here a total or resultant displacement of which δu_1 and δu_2 are the components. Thus the limit of $\delta P/\delta s$ is to be regarded as a *total*, not a partial differential coefficient, and must be symbolized by dP/ds. Similarly, the limits of $\delta u_1/\delta s$ and $\delta u_2/\delta s$ are total differential coefficients; for s is not one of a set of independent variables of which the others remain constant while s varies. Thus the limit of (23 : 5) must be written

$$\frac{dP}{ds} = \frac{\partial P}{\partial u_1}\frac{du_1}{ds} + \frac{\partial P}{\partial u_2}\frac{du_2}{ds}$$

Similarly, for four dimensions we have

$$\frac{dP}{ds} = \sum_a \frac{\partial P}{\partial u_a}\frac{du_a}{ds} \quad [a = 1, 2, 3, 4] \qquad (23:6)$$

(iv) We reached (23 : 4) by argument from a special case, but the result, as we have already pointed out, is quite

general. For instance, if for P we substitute one of the coordinates of the V-system, say v_n, we have

$$\frac{\partial v_n}{\partial v_m} = \sum_a \frac{\partial v_n}{\partial u_a} \frac{\partial u_a}{\partial v_m}$$

Now two cases may arise : (a) v_n and v_m may represent the same coordinate ; in that case the value of $\partial v_n / \partial v_m$ is unity. (b) They may represent different coordinates of the same system ; in that case they are independent of one another. And since, by hypothesis, as v_m changes the independent v_n remains constant, the partial differential coefficient is now zero. Hence the important results :

$$\begin{aligned} \sum_a \frac{\partial v_n}{\partial u_a} \frac{\partial u_a}{\partial v_m} &= 1 \qquad (m = n) \\ &= 0 \qquad (m \neq n) \end{aligned} \qquad (23:7)$$

§ 24. *Variations. Geodesics.*—Fig. 4 is part of the graph

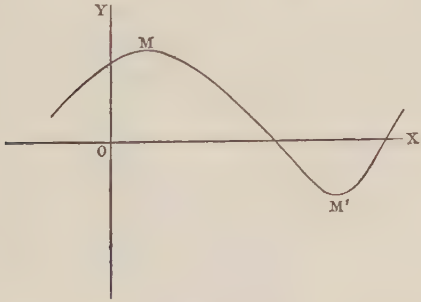

FIG. 4.

of a function $y = f(x)$ which has a maximum and a minimum value at the points M and M' respectively. These facts are sometimes expressed by saying that at M and M' the value of y is " stationary "—the meaning

being that for a small change δx in x there is no change at all in y, or $\delta y = 0$. This way of looking at the matter is, of course, only approximately true; but it is sometimes convenient and we use it in what follows.

(i) In fig. 5 let a pair of points A, B be joined by an

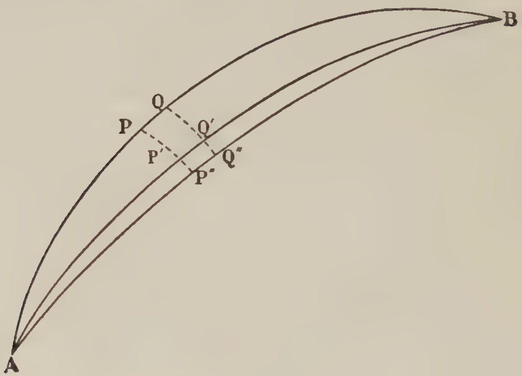

FIG. 5.

indefinite number of curves. In the case of one of these, namely APB, let the distance along the curve from A to any point P be denoted by s, and let the whole curve be divided into short portions δs of which PQ is a specimen. Further, let each of the other curves be divided up in correspondence with the divisions of APB, so that the elements $P'Q'$, $P''Q''$, etc., correspond to PQ. In the case of one of these curves, say $AP'B$, let the length of an element be $\delta s'$; then we can put

$$\delta s' = w\delta s$$

where w is the ratio of the lengths of the corresponding elements.

The ratio w is, of course, not necessarily constant

but will as a rule vary in value from element to element along the line. Nevertheless, the *series* of values of w will be a definite one for a particular curve, so that each member of the group of curves will be characterized by its own series of w's. The length of any curve will be

$$\int_A^B ds' = \int_A^B w ds \qquad (24:1)$$

As we pass from one curve to another the length measured by the integral (24 : 1) will in general increase or decrease in accordance with changes in the series of w's, but where it reaches a minimum or a maximum * it will be stationary. In accordance with the explanation given above, we then have

$$\delta\int_A^B w ds = 0 \qquad (24:2)$$

In technical terms, the "variation" in the integral, as we pass from one series of w's to one very little different from it, is zero when the integral's value is stationary.

Now the length of our curve is the limit, as n increases indefinitely, of the sum

$$w_1\delta s_1 + w_2\delta s_2 + w_3\delta s_3 + \;\ldots\ldots + w_n\delta s_n$$

while the length of a neighbouring curve in which the series of w's is almost the same as in this may be written as the limit of

$$(w_1 + \delta w_1)\,\delta s_1 + (w_2 + \delta w_2)\,\delta s_2 + \ldots\ldots + (w_n + \delta w_n)\,\delta s_n$$

* In ordinary space there could be no maximum, but if a maximum were possible the argument would apply. In the case of four-dimensional space-time we have a maximum as indicated in § 22. There is also a minimum, the length of the world line representing the movement of a light-ray.

Thus we have for the difference in length of two neighbouring curves

$$\delta\int_A^B w ds = \underset{n\to\infty}{Lt}\Sigma\left[\delta w_1\delta s_1 + \delta w_2\delta s_2 + \ldots\ldots + \delta w_n\delta s_n\right]$$

$$= \int_A^B \delta w ds \qquad (24:3)$$

whence it follows that the curve of stationary length is the one corresponding to the equation

$$\int_A^B \delta w ds = 0 \qquad (24:4)$$

Note that a curve of stationary length between two given points is called a " geodesic ". As examples of geodesics we have : (i) the straight line joining two points A and B in ordinary space ; (ii) the shortest route joining A and B on a curved surface ; (iii) the " uniform world line " in four-dimensional space-time (§ 22).

(ii) In the application of the preceding argument in Chapter VII we shall require the following theorem :

$$\delta\left(\frac{dy}{dx}\right) = \frac{d}{dx}\overset{\delta}{(y)}$$

In words: the difference between the differential coefficients of y for any two near values of y is equal to the differential coefficient of the difference between those values of y. Let y' be a value near to y, so that $y' - y$ is δy ; and note in this connexion that if y' and y are *both* slightly increased, the change produced in their difference may be expressed in three equivalent ways :

$$\delta(y' - y) = \delta y' - \delta y = \delta(\delta y)$$

Now, by definition,

$$\frac{dy'}{dx} = Lt\,\frac{\delta y'}{\delta x} \qquad \frac{dy}{dx} = Lt\,\frac{\delta y}{\delta x}$$

whence

$$\delta\left(\frac{dy}{dx}\right) = \frac{dy'}{dx} - \frac{dy}{dx} = Lt\,\frac{\delta y' - \delta y}{\delta x} = Lt\,\frac{\delta(\delta y)}{\delta x} = \frac{d}{dx}(\delta y) \qquad (24:5)$$

§ 25. *Determinants.*—By definition, the determinant

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1(b_2c_3 - b_3c_2) - a_2(b_1c_3 - b_3c_1) + a_3(b_1c_2 - b_2c_1) \qquad (25:1)$$

The expressions in the brackets are called the "minors" of the elements outside the brackets. They can themselves be written as determinants, namely:

$$\begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} \qquad \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} \qquad \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix}$$

Note that the minor of a term in the larger determinant is obtained by missing out the row and column in which it appears, and that in reconstructing the larger determinant the terms must be taken alternately positive and negative.

An alternative expression for the determinant can be obtained by multiplying the terms of any other row by their minors and adding the products. For instance, if we work with the second row the value of the determinant becomes

$$-b_1(a_2c_3 - a_3c_2) + b_2(a_1c_3 - a_3c_1) - b_3(a_1c_2 - a_2c_1) \qquad (25:2)$$

which is obviously in agreement with the former expression.

Note from this case the necessity of alternating the

signs vertically to find the sign of the first term as well as horizontally in working out the development.

It can easily be shown that if the rows are converted into columns and the columns into rows, the value of the determinant is unchanged.

(i) Let the second and third rows be interchanged; then since in (25 : 1) the b's and c's change places, it is obvious that the sign of the determinant is reversed. From (25 : 2) it is clear that the same thing would happen if the third and first rows were interchanged. The theorem can be demonstrated similarly for interchange of the second and third rows.

(ii) Let two rows be made identical. Then interchange of these rows can make no difference to the value of the determinant. But by (i) its sign is reversed. It follows that a determinant with two identical rows (or columns) must have zero value.

(iii) Let the minors of the terms in the first row be symbolized by A_1, $-A_2$, A_3; then the value of the determinant is

$$a_1A_1 + a_2A_2 + a_3A_3$$

Now consider the sums

$$b_1A_1 + b_2A_2 + b_3A_3$$

and

$$c_1A_1 + c_2A_2 + c_3A_3$$

The first sum shows what the value of the determinant would be if the first row were identical with the second, the second sum what it would be if the first row were identical with the third. But in each of these cases the value would, by (ii), be zero. We conclude that if we work along any row of the determinant, multiplying the terms of that row by the minors of *another* row with

their proper signs, the sum of the products will in each case be zero.

(iv) The determinants used in the theory of relativity are practically always determinants of 16 elements. Let the nth element in the mth row be written a_{mn} and its minor be written $(-)^{m+n}A_{mn}$. Then we have for the value of the determinant

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix} = a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13} + a_{14}A_{14}$$

It can be shown without difficulty that the theorems proved in (i), (ii) and (iii) also hold good for determinants of 16 terms.

(v) Let the value of the determinant be called a and let*

$$a^{mn} = A_{mn}/a \qquad (25:3)$$

Now consider the sum

$$\sum_n a_{mn} \cdot a^{mn}$$

The expression implies that while m retains a particular value, n assumes in succession the values 1, 2, 3, 4. Suppose $m = 2$; then we are directed to work along the second row and to obtain the sum

$$a_{21} \cdot a^{21} + a_{22} \cdot a^{22} + a_{23} \cdot a^{23} + a_{24} \cdot a^{24}$$

But this is, by definition, equal to

$$\frac{a_{21} \cdot A_{21} + a_{22} \cdot A_{22} + a_{23} \cdot A_{23} + a_{24} \cdot A_{24}}{a} = \frac{a}{a} = 1$$

The same result would be obtained if m had any other of its possible values. In general, then,

$$\sum_n a_{mn} \cdot a^{mn} = 1$$

* Observe that *amn* does *not* mean the *mn*th *power* of *a*. The justification for the symbolism will appear in Ch. X.

Consider next the sum

$$\sum_n a_{mn} \cdot a^{pn} = \frac{a_{m1} \cdot A_{p1} + a_{m2}A_{p2} + a_{m3}A_{p3} + a_{m4}A_{p4}}{a}$$

Now by (iii), if p and m are different numbers, the numerator is zero; for it is the sum of the products obtained by working along the mth row and multiplying each term of that row by the minor of the corresponding term in the pth row.

We have thus demonstrated the extremely important theorem

$$\begin{aligned} \sum_n a_{mn} \cdot a^{pn} &= 1 \qquad (m = p) \\ &= 0 \qquad (m \neq p) \end{aligned} \qquad (25:4)$$

§ 26. *Polar Coordinates.*—It will sometimes be convenient to express (19 : 6) in terms of polar coordinates. In fig. 6 let the position of P be fixed by specifying r,

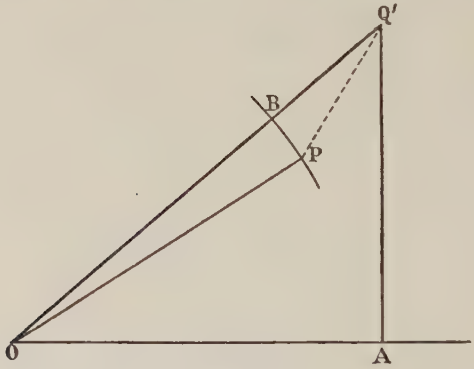

Fig. 6.

which is the length of OP, and the angle $AOP = \theta$. The coordinates of a point Q', near P and in the plane AOP, would then be $r + \delta r$ and $\theta + \delta\theta$; where $BQ' = \delta r$ and $\delta\theta$ is the small angle POB.

Imagine another point Q a very short distance from Q' up the perpendicular to the paper. From Q' draw $Q'A$ at right angles to OA, and let the small angle $Q'AQ = \delta\phi$. Then

$$Q'Q = Q'A\,\delta\phi = OQ' \sin Q'OA\,.\,\delta\phi = (r + \delta r) \sin (\theta + \delta\theta)\,.\,\delta\phi$$
$$= r \sin \theta\,.\,\delta\phi$$

to the first order of small quantities.

Now let P and Q be the points where two near event-particles occur. Then we have for the distance PQ

$$PQ^2 = PQ'^2 + QQ'^2 = Q'B^2 + BP^2 + QQ'^2$$
$$= \delta r^2 + r^2\delta\theta^2 + r^2 \sin^2 \theta\delta\phi^2$$

Thus the polar formula corresponding to (19 : 6) is

$$\delta s^2 = -\delta r^2 - r^2\delta\theta^2 - r^2 \sin^2 \theta\delta\phi^2 + c^2\delta t^2 \qquad (26:1)$$

Next let Q be *any* distance up the perpendicular $Q'Q$, and let the angle $AOQ = \theta$. Then

$$OA = OQ \cos \theta,\ QA = OQ \sin \theta$$
$$Q'A = QA \cos \phi,\ QQ' = QA \sin \phi$$

It follows that if $OQ = r$, $OA = x$, $AQ' = y$ and $QQ' = z$, then

$$x = r \cos \theta,\ y = r \sin \theta \cos \phi,\ z = r \sin \theta \sin \phi \quad (26:2)$$

Whence the following relations between the differentials:

$$\delta x = \cos \theta\, \delta r - r \sin \theta\, \delta\theta$$
$$\delta y = \sin \theta \cos \phi\, \delta r + r \cos \theta \cos \phi\, \delta\theta - r \sin \theta \sin \phi\, \delta\phi$$
$$\delta z = \sin \theta \sin \phi\, \delta r + r \cos \theta \sin \phi\, \delta\theta + r \sin \theta \cos \phi\, \delta\phi$$
$$(26:3)$$

CHAPTER VII

THE GEODESIC LAW OF MOTION

§ 27. THE restricted theory of relativity is concerned only with systems whose relative motion is uniform; in the general theory the relative motion of the systems may be of any kind. The firstfruits of widening the outlook have already been gathered in Chapter III. We saw there that the space-regions considered in theoretical physics include: (i) regions in which free particles would move, always and everywhere, with uniform rectilinear motion in accordance with the law of Galileo; (ii) uniform fields of force in which they would move, always and everywhere, with a constant acceleration a; (iii) permanent gravitational fields in which the acceleration of a particle is always the same at the same point, but varies from place to place. We may for completeness add that there are (iv) regions in which the acceleration varies from time to time as well as (possibly) from place to place—as in cases of wave-transmission. But these last, though of immense importance, lie outside our purview. We further saw that a region which is of type (ii) from the standpoint of a particular coordinate system S can always be reduced to type (i) by being referred to a coordinate system S' moving with acceleration $-a$ with regard to S. Thus in the theory of relativity types (i) and (ii) are merely subordinate divisions of one type, the Galilean. On the other hand we found that a region which belongs to type (iii) from the standpoint of any one system belongs to

the same type for all systems. In accordance with the Principle of Equivalence, a small region of a permanent gravitational field may be Galilean from the standpoint of a suitably chosen S'-system, but there is no standpoint from which the whole field is Galilean.

An extension of the spatial analogies of Chapter V may be used to express the last point. Just as a plane, which is a surface of zero curvature, may be a tangent to a curved surface and may be conceived as coinciding with a small area about the point of contact, so a Galilean space-time continuum may be thought of as tangential to a non-Galilean continuum and as possessing a small four-dimensional region in common with it. The reader will find here the meaning of the statement that in a permanent gravitational field space-time possesses curvature; a contrast is implied with Galilean space-time in which all motion is uniform in the sense explained in § 22. The analogy also suggests that there may be types of motion not amenable to the principle of equivalence. For just as mathematicians have shown that there are tangentless curves, so there may be regions of space-time which cannot be regarded as coincident with a Galilean continuum even as regards their smallest elements. We shall, however, proceed on the assumption that the principle of equivalence is applicable in all cases we have to deal with, just as it is assumed in elementary mathematics that all curves have tangents.

Our task in the present chapter is to seek a law of motion for particles which will hold good alike in Galilean and non-Galilean space-time. Newton's second law* claimed

* "Change of motion is proportional to the impressed force, and takes place in the direction of the straight line in which the force acts."

universal validity, but has been rendered obsolete (except as an indispensable approximation) by the principle of relativity. Relieved of the metaphysical and superfluous notion of "force", it asserts, in effect, that with every point of a permanent gravitational field there is associated a definite acceleration, and that if a particle should find its way to that point its motion would there exhibit that acceleration. In analytical terms we may say that there is at every point a potential P (§ 23 (ii)), such that the acceleration is given by the equations

$$\frac{d^2x}{dt^2} = \frac{\partial P}{\partial x}, \quad \frac{d^2y}{dt^2} = \frac{\partial P}{\partial y}, \quad \frac{d^2z}{dt^2} = \frac{\partial P}{\partial z} \qquad (27:1)$$

But inasmuch as the space and time measures of different systems differ with their relative motion, and no system can claim a prerogative vote, it is impossible to specify such a set of invariant accelerations as Newton's theory contemplates.* Thus the law necessarily fails. If, then, it be asked how the great fabric of modern physics and astronomy could be built upon it, the answer is, in the first place, that the velocities and accelerations of the material bodies studied by physicists and astronomers were relatively very small, and in the second place that the relative velocities of the alternative systems of reference which it was necessary to take into account were so small compared with c that their time-flows could be regarded as identical. If the ratio v/c is negligible; $t' = t$ in the Lorentz transformation, and it then follows

* Remember that even in a permanent gravitational field, such as the sun's, though the motion of S' cannot destroy the set of accelerations observed by S, it nevertheless changes all their values.

that $x' = x - vt$. Now from this relation we deduce that

$$\frac{dx'}{dt'} = \frac{dx'}{dt} = \frac{dx}{dt} - v$$

that is, that velocities observed in the S'-system are different from those observed in the S-system. But when we differentiate a second time we have, since v is constant,

$$\frac{d^2x'}{dt'^2} = \frac{d}{dt}\left(\frac{dx'}{dt'}\right) = \frac{d^2x}{dt^2}$$

that is, accelerations observed in the two systems are identical.

§ 28. In § 19 it was shown that the formula

$$\delta s^2 = -\delta x^2 - \delta y^2 - \delta z^2 + c^2\delta t^2 \qquad (28:1)$$

or its polar equivalent (26 : 1)

$$\delta s^2 = -\delta r^2 - r^2\delta\theta^2 - r^2 \sin^2\theta\, \delta\phi^2 + c^2\delta t^2 \qquad (28:2)$$

holds good for all conceivable rectangular systems in any state of relative motion. It is, however, of the utmost importance to observe that the argument rested upon the assumption that space-time is uniform or " homaloidal ". This assumption is implied by the very attempt to lay down universal rules for the transformation of coordinates from one system to another; for if different regions of space-time differed in character universal rules would be impossible. And the reader will recollect that it was explicitly made in the argument leading to the Lorentz formulæ. But the deduction of (28 : 1) was based on those formulæ and cannot be expected, therefore, to hold good where the assumption of uniformity cannot be made.

Now as we saw in § 10, the corner-stone of Einstein's

theory of gravitation is that space-time is *not* uniform, and that its varying intrinsic character accounts for the varying acceleration of a particle as it moves through the gravitational field. It follows that (28 : 1) will hold good only in Galilean regions, and that elsewhere its form will be changed.

In view of what is to come it will be profitable to inquire what kinds of modification are to be expected. Let us make first the fantastic supposition that the "attracting" matter of the universe constitutes a circular disc with a radius of (say) a million million kilometres. Take the plane of the surface for the yz-plane and the axis of the disc for the x-axis, and let attention be confined to a cylinder of space having a radius of a couple of hundred million kilometres about the x-axis and stretching the same distance from the disc's surface. Within such a region we are, following Einstein, to expect the properties of space-time to be modified, so to speak, in layers parallel to the disc*—the greatest modification being in the nearest layers and the deviations from Galilean uniformity fading out as we recede into the depths of space. In such a case we should not expect (28 : 1) to be changed by the introduction of new terms, such as $\delta x.\delta y$ or $\delta z.\delta t$. For if we take a certain value of δx and associate with it in succession equal but opposite values of δy, the effects upon δs^2 would be contradictory. But since the field, in the limited region we are studying, is uniform at a constant distance from the disc, such a lack of symmetry in the expression for δs^2 is ruled out. Again, the introduction of terms such as $\delta z.\delta t$ would imply a difference between the properties of space-time at a given place at

* As the edge is approached the layers will cease to be parallel.

times δt before and after a given epoch. But there is no reason why the field should change with time. Thus the terms in the expression for δs^2 will remain those of (28 : 1), the only differences being in their coefficients. In short we shall have an expression of the type

$$\delta s^2 = -A\delta x^2 - B\delta y^2 - B\delta z^2 + Cc^2\delta t^2 \tag{28 : 3}$$

the coefficients of δy^2 and δz^2 being made identical in view of the symmetry of the disc about its axis.

Now it is to be observed that A, B and C cannot be constants. If they were, there would still be no difference between the properties of space-time at different distances from the disc. In fact, by assuming new units of measurement X, Y, Z, T, such that $\delta X/\delta x = \sqrt{A}$, etc., we should return to (28 : 1) in the form

$$\delta s^2 = -\delta X^2 - \delta Y^2 - \delta Z^1 + c^2\delta T$$

It is clear, therefore, that the introduced coefficients must be functions of x of such a nature that they all approach unity as x increases.

Next take the more actual case of the sun, regarded as a dense particle in an otherwise empty universe. Here it is natural to employ the polar formula (28 : 2). Making the inevitable assumption that space-time is symmetrical in its properties about any radial line drawn from the sun, we see as before that no new terms will be introduced which involve the product of any two of the infinitesimals δr, $\delta\theta$, etc. Thus (28 : 2) will be modified into

$$\delta s^2 = -A\delta r^2 - Br^2\delta\theta^2 - Br^2\sin^2\theta\,\delta\phi^2 + Cc^2\delta t^2 \tag{28 : 4}$$

where the coefficients A, B and C are functions of r

only, and approach unity as r increases.* It is clear that in both examples the amounts by which, in any place, the coefficients differ from unity measure the amount by which space-time deviates in that place from uniformity.

§ 29. Formulæ (28 : 3, 4) were taken to express the modification of space-time in particular instances where the presence of symmetry in the nature of the field precluded a more drastic departure from (28 : 1). In the most general case we must, analogously, assume that δs^2 is determined by the complete quadratic formula

$$\begin{aligned}\delta s^2 = g_{11}\delta x^2 + g_{22}\delta y^2 + g_{33}\delta z^2 + g_{44}c^2\delta t^2 \\ + 2g_{12}\delta x\delta y + 2g_{13}\delta x\delta z + 2g_{14}c\delta x\delta t \\ + 2g_{23}\delta y\delta z + 2g_{24}c\delta y\delta t + 2g_{34}c\delta z\delta t\end{aligned} \tag{29 : 1}$$

where the coefficients (they will often be referred to as "the g's") are functions of x, y, z and t. But though the separation between two near event-particles is now measured in a more complicated way the general argument of § 19 still holds good. It is an *intrinsic* property of the space-time region and quite independent of particular systems of reference and their modes of motion. In other words, if δs is the separation between a definite pair of event-particles, its value will come out as exactly the same number in whatever system it is measured. But the relation between δs and the differentials of the coordinates is one which varies from region to region in accordance with the manner in which those regions diverge from the uniformity of Galilean space-time.

It will generally be convenient to replace x, y, z, ct by

* Since $r\delta\theta$ and $r\sin\theta\,\delta\phi$ are elements perpendicular to one another and to the radius-vector, symmetry requires that the coefficients (B) should be identical.

v_1, v_2, v_3, v_4, and to consider the 10 terms of (29 : 1) as 16, which may be set out as follows :

$$\begin{aligned}\delta s^2 = {} & g_{11}\delta v_1\delta v_1 + g_{12}\delta v_1\delta v_2 + g_{13}\delta v_1\delta v_3 + g_{14}\delta v_1\delta v_4 \\ & + g_{21}\delta v_2\delta v_1 + g_{22}\delta v_2\delta v_2 + g_{23}\delta v_2\delta v_3 + g_{24}\delta v_2\delta v_4 \\ & + g_{31}\delta v_3\delta v_1 + g_{32}\delta v_3\delta v_2 + g_{33}\delta v_3\delta v_3 + g_{34}\delta v_3\delta v_4 \\ & + g_{41}\delta v_4\delta v_1 + g_{42}\delta v_4\delta v_2 + g_{43}\delta v_4\delta v_3 + g_{44}\delta v_4\delta v_4\end{aligned} \tag{29 : 2}$$

On comparison with (29 : 1) it will be seen that coefficients of the forms g_{ab} and g_{ba} are always the same ; the distinction between them is made only with a view to obtaining the symmetrical arrangement of (29 : 2).

§ 30. The arrangement shown in (29 : 2) is particularly useful because it suggests that the g's may be regarded as leading to a determinant :

$$\begin{vmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & g_{34} \\ g_{41} & g_{42} & g_{43} & g_{44} \end{vmatrix} = g \tag{30 : 1}$$

In accordance with the definition given in (25 : 3) the symbol g^{mn} means the minor of g_{mn} divided by the value of the determinant, g. Quantities of this type play an extremely important part in the theory of relativity.

The determinants for the g's of Galilean space-time are

$$\begin{vmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & c^2 \end{vmatrix} = -c^2, \quad (30 : 2) \qquad \begin{vmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 \end{vmatrix} = -1 \quad (30 : 3)$$

and

$$\begin{vmatrix} -1 & 0 & 0 & 0 \\ 0 & -r^2 & 0 & 0 \\ 0 & 0 & -r^2\sin^2\theta & 0 \\ 0 & 0 & 0 & c^2 \end{vmatrix} = -c^2r^4\sin^2\theta \tag{30 : 4}$$

The first corresponds to (28 : 1), the second to the form

$$\delta s^2 = - \delta v_1^2 - \delta v_2^2 - \delta v_3^2 + \delta v_4^2 \qquad (30:5)$$

the third to the polar formula (28 : 2).

§ 31. We are now in a position to follow Einstein's deduction of the universal law of motion for a particle. The argument has two stages:

(i) Let δs be the separation of two near point-instants on the world line which represents the track of a particle in space-time. Then, as we have seen, its value at a particular place on the line is entirely independent of the coordinate system in which it may be measured. It follows that the integral

$$\int_A^B ds$$

between any two points-events A and B which are situated on the world line is also the same for every coordinate system. In particular, if the world line joining A and B is such that the value of the integral is "stationary"—i.e. if it is a geodesic (§ 24)—for any one system, then it is a geodesic for all systems.

(ii) Now let the moving particle be accompanied by an observer S in the manner described in § 9. Then by the principle of equivalence the immediate neighbourhood will constantly be viewed by S as a Galilean region through which the particle is moving uniformly in a straight line. Since this applies all along the track, it applies to the whole; hence (§ 22) the world line which presents the history of the movement, from S's standpoint, in four-dimensional space is a geodesic. But as shown in (i) this implies that the world line will be a geodesic in every possible coordinate system.

The law of motion which Einstein offers as a substitute for Newton's may, then, be formulated thus: The world line of a free particle is always a geodesic. This law possesses the character of universality which Newton's lacks; for it holds good for every possible system of coordinates to which the particle's history might be referred.

§ 32. It remains to discover the conditions for a geodesic world line between two given event-particles A and B.

As in fig. 5, § 24, let APB be any world line running through A and B, and in order that the treatment may be completely general, assume (29 : 2) as the formula for δs. This may be expressed concisely in the form

$$\delta s^2 = \sum_m \sum_n g_{mn} \delta v_m \delta v_n \qquad [m, n = 1, 2, 3, 4] \qquad (32:1)$$

Simultaneous multiplication and division of (32 : 1) by δs^2 converts it into the form

$$\delta s^2 = \left[\sum_m \sum_n g_{mn} \frac{\delta v_m}{\delta s} \frac{\delta v_n}{\delta s} \right] \delta s^2 \qquad (32:2)$$

from which it is obvious that in the limit

$$\sum_m \sum_n g_{mn} \frac{dv_m}{ds} \frac{dv_n}{ds} = 1 \qquad (32:3)$$

Let $AP'B$ be a second world line through A and B, and let its separation-element be given by

$$\delta s'^2 = \sum_m \sum_n g'_{mn} \delta v'_m \delta v'_n \qquad (32:4)$$

accents being employed here merely to distinguish the

quantities from those referred to in (32 : 1). As before it follows that

$$\sum_m \sum_n g'_{mn} \frac{dv'_m}{ds'} \frac{dv'_n}{ds'} = 1 \qquad (32:5)$$

Now suppose, as explained in § 24, that $\delta s' = w\delta s$; then we deduce from (32 : 5) that

$$w^2 = \sum_m \sum_n g'_{mn} \frac{dv'_m}{ds} \frac{dv'_n}{ds} \qquad (32:6)$$

In accordance with § 24 the condition that $AP'B$ is a geodesic is that

$$\int_A^B \delta w ds = 0 \qquad (32:7)$$

Now let the position of $AP'B$ differ only slightly from that of APB. Then since $w = 1$ along APB, its values along $AP'B$ may be written $(1 + \delta w)$. At the same time we may put

$$g'_{mn} = g_{mn} + \delta g_{mn}$$

$$\frac{dv'_m}{ds} = \frac{dv_m}{ds} + \delta\left(\frac{dv_m}{ds}\right), \qquad \frac{dv'_n}{ds} = \frac{dv_n}{ds} + \delta\left(\frac{dv_n}{ds}\right)$$

Thus (32 : 6) becomes

$$(1 + \delta w)^2 = \sum_m \sum_n (g_{mn} + \delta g_{mn})\left\{\frac{dv_m}{ds} + \delta\left(\frac{dv_m}{ds}\right)\right\}\left\{\frac{dv_n}{ds} + \delta\left(\frac{dv_n}{ds}\right)\right\}$$

whence (neglecting the square and products of the small quantities) we have

$$\begin{aligned}(1 + 2\delta w) = {} & \sum_m \sum_n g_{mn} \frac{dv_m}{ds} \frac{dv_n}{ds} \\ & + \sum_m \sum_n \delta g_{mn} \frac{dv_m}{ds} \frac{dv_n}{ds} \\ & + \sum_m \sum_n g_{mn} \left\{\frac{dv_n}{ds} \delta\left(\frac{dv_m}{ds}\right) + \frac{dv_m}{ds} \delta\left(\frac{dv_n}{ds}\right)\right\}\end{aligned}$$

In virtue of (32 : 3) this equality reduces to

$$2\delta w = \sum_m \sum_n \left[\delta g_{mn} \frac{dv_m}{ds} \frac{dv_n}{ds} + g_{mn} \left\{ \frac{dv_n}{ds} \delta \left(\frac{dv_m}{ds} \right) + \frac{dv_m}{ds} \delta \left(\frac{dv_n}{ds} \right) \right\} \right]$$

$$= \sum_m \sum_n \left[\delta g_{mn} \frac{dv_m}{ds} \frac{dv_n}{ds} + 2 g_{mn} \left\{ \frac{dv_m}{ds} \delta \left(\frac{dv_n}{ds} \right) \right\} \right] \qquad (32:8)$$

The last step is justified by the observation that if, in obedience to the double summation sign, the 16 values of each of the two expressions in the curled bracket were written out, they would be identical, and would differ only in order. For instance, when $m = 2$, $n = 3$,

$$\frac{dv_n}{ds} \delta \left(\frac{dv_m}{ds} \right) = \frac{dv_3}{ds} \delta \left(\frac{dv_2}{ds} \right), \quad \frac{dv_m}{ds} \delta \left(\frac{dv_n}{ds} \right) = \frac{dv_2}{ds} \delta \left(\frac{dv_3}{ds} \right)$$

while when $m = 3$, $n = 2$,

$$\frac{dv_n}{ds} \delta \left(\frac{dv_m}{ds} \right) = \frac{dv_2}{ds} \delta \left(\frac{dv_3}{ds} \right), \quad \frac{dv_m}{ds} \delta \left(\frac{dv_n}{ds} \right) = \frac{dv_3}{ds} \delta \left(\frac{dv_2}{ds} \right)$$

Again, by an argument analogous to that of § 23 (i),

$$\delta\, g_{mn} = \sum_a \frac{\partial g_{mn}}{\partial v_a} \delta v_a \qquad [a = 1, 2, 3, 4]$$

and by (24 : 5)

$$\delta \left(\frac{dv_n}{ds} \right) = \frac{d}{ds} (\delta v_n)$$

Thus (32 : 8) becomes

$$\delta w = \sum_m \sum_n \left[\frac{1}{2} \sum_a \frac{\partial g_{mn}}{\partial v_a} \frac{dv_m}{ds} \frac{dv_n}{ds} \delta v_a + g_{mn} \frac{dv_m}{ds} \frac{d}{ds} (\delta v_n) \right]$$

$$= \qquad \tfrac{1}{2} P \qquad + Q$$

and $$\int_A^B \delta w\, ds = \tfrac{1}{2} \int_A^B P\, ds + \int_A^B Q\, ds \qquad (32:9)$$

Treating the second integral on the right in accordance with the rule for integration by parts:

$$\int y dx = (xy) - \int x dy$$

we have

$$\int_A^B Q ds = \sum_m \sum_n \left[\left[g_{mn} \frac{dv_m}{ds} \delta v_n \right]_A^B - \int_A^B \delta v_n \frac{d}{ds} \left(g_{mn} \frac{dv_m}{ds} \right) ds \right]$$

$$= 0 - \sum_m \sum_n \int_A^B \delta v_n \frac{d}{ds} \left(g_{mn} \frac{dv_m}{ds} \right) ds \qquad (32:10)$$

For since all the world lines under consideration pass through A and B, δv_n (i.e. the change in the v_n-coordinate in passing from one line to another) is zero at both limits.

Now, $$\frac{d}{ds}\left(g_{mn} \frac{dv_m}{ds} \right) = g_{mn} \frac{d^2 v_m}{ds^2} + \frac{dv_m}{ds} \frac{d}{ds} (g_{mn})$$

$$= g_{mn} \frac{d^2 v_m}{ds^2} + \sum_a \frac{\partial g_{mn}}{\partial v_a} \frac{dv_a}{ds} \frac{dv_m}{ds} \qquad (32:11)$$

by (23 : 6). Thus (32 : 9) may be written, with reversed signs,

$$-\int_A^B \delta w ds = \int_A^B \sum_m \sum_n \left[g_{mn} \frac{d^2 v_m}{ds^2} \delta v_n + \sum_a \left\{ \frac{\partial g_{mn}}{\partial v_a} \frac{dv_a}{ds} \frac{dv_m}{ds} \delta v_n \right. \right.$$

$$\left. \left. - \tfrac{1}{2} \frac{\partial g_{mn}}{\partial v_a} \frac{dv_m}{ds} \frac{dv_n}{ds} \delta v_a \right\} \right] ds \qquad (32:12)$$

The next step is to replace δv_n by δv_a or *vice versa*. We have

$$\sum_m \sum_n g_{mn} \frac{d^2 v_m}{ds^2} \delta v_n = \sum_m \sum_a g_{ma} \frac{d^2 v_m}{ds^2} \delta v_a$$

and

$$\sum_m \sum_n \sum_a \frac{\partial g_{mn}}{\partial v_a} \frac{dv_m}{ds} \frac{dv_a}{ds} \delta v_n = \sum_m \sum_a \sum_n \frac{\partial g_{ma}}{\partial v_n} \frac{dv_m}{ds} \frac{dv_n}{ds} \delta v_a$$

$$= \sum_m \sum_n \sum_a \left[\tfrac{1}{2} \frac{\partial g_{ma}}{\partial v_n} + \tfrac{1}{2} \frac{\partial g_{na}}{\partial v_m} \right] \frac{dv_m}{ds} \frac{dv_n}{ds} \delta v_a$$

by a reversal of the argument used to justify (32 : 8).

Hence, if we put $[mn, a]$ for $\frac{1}{2}\left\{\frac{\partial g_{ma}}{\partial v_n} + \frac{\partial g_{na}}{\partial v_m} - \frac{\partial g_{mn}}{\partial v_a}\right\}$, equation (32 : 12) becomes

$$-\int_A^B \delta w\, ds = \int_A^B \left[\sum_m \sum_n \sum_a \left\{ g_{ma} \frac{d^2 v_m}{ds^2} + [mn, a] \frac{dv_m}{ds} \frac{dv_n}{ds} \right\} \delta v_a \right] ds = 0 \qquad (32:13)$$

Now δv_a represents a completely arbitrary change in the value of a coordinate, and is therefore susceptible of an infinite variety of values. It follows that if the integral is to be zero

$$\sum_m \sum_a g_{ma} \frac{d^2 v_m}{ds^2} + \sum_m \sum_n \sum_a [mn, a] \frac{dv_m}{ds} \frac{dv_n}{ds} = 0 \qquad (32:14)$$

In a sense this solves our problem ; but the solution is of no practical use because, on account of the summation with regard to m, it involves all the coordinates equally. We can, however, isolate any one of them—say v_p, where p is a particular one of the four numbers 1, 2, 3, 4—by multiplying (32 : 14) by g^{pa}, *i.e.* the quantity whose definition and properties were given in § 25 (v) and § 30. We then have

$$\sum_m \left[\sum_a g_{ma} \cdot g^{pa} \right] \frac{d^2 v_m}{ds^2} = \frac{d^2 v_p}{ds^2} + 0 + 0 + 0$$

For by (25 : 4) $\sum_a g_{ma} \cdot g^{pa}$ is zero unless $m = p$, and is then unity.

Thus the condition that the world line may be a geodesic becomes

$$\frac{d^2v_p}{ds^2} + \sum_m \sum_n \left[\sum_a g^{pa} [mn, a] \right] \frac{dv_m}{ds} \frac{dv_n}{ds} = 0$$

which is usually expressed in the form

$$\frac{d^2v_p}{ds^2} + \sum_m \sum_n \{mn, p\} \frac{dv_m}{ds} \frac{dv_n}{ds} = 0 \qquad (32:15)$$

This equation must, of course, be satisfied for each of the four values of p.

The expressions $[mn, a]$ and $\{mn, p\}$ are called the first and second "3-index symbols of Christoffel". Since we shall make much use of them, the reader should take careful note of their definitions:

$$[mn, a] \equiv \tfrac{1}{2}\left(\frac{\partial g_{ma}}{\partial v_n} + \frac{\partial g_{na}}{\partial v_m} - \frac{\partial g_{mn}}{\partial v_a}\right) \qquad (32:16)$$

$$\{mn, p\} \equiv \sum_a g^{pa} [mn, a]$$

$$= \sum_a \tfrac{1}{2} g^{pa} \left(\frac{\partial g_{ma}}{\partial v_n} + \frac{\partial g_{na}}{\partial v_m} - \frac{\partial g_{mn}}{\partial v_a}\right) \qquad (32:17)$$

§ 33. Equation (32 : 15) is Einstein's universal law of motion for particles. It looks strangely different from Newton's, yet if it is true the two must be in harmony. The criticism of Newton's law is that it is, so to speak, parochial instead of being worldwide in its scope. In a word, it was the discovery of a very great genius whose experience was limited to velocities small compared with c, and to space-time with properties scarcely differing from the Galilean. If we submit Einstein's law to the same limitations, the two ought to become indistinguishable.

The condition that space-time is to be very nearly Galilean is that the g's are very little different from those set out in (30 : 2, 3 or 4); or that δs^2 is expressed with approximate accuracy by

$$\begin{aligned}\delta s^2 &= -(\delta x^2 + \delta y^2 + \delta z^2) + c^2\delta t^2\\ &= -\quad r^2 + c^2\delta t^2\end{aligned}$$

and the condition that the particle's velocity is small compared with that of light is that δr^2 is negligible compared with $c^2\delta t^2$. In that case δs would be nearly equal to $c\delta t$, and the following approximate equivalences would hold good:

$$\frac{dx}{ds} = \frac{1}{c}\frac{dx}{dt} = 0, \quad \frac{dy}{ds} = 0, \quad \frac{dz}{ds} = 0, \quad \frac{dt}{ds} = \frac{1}{c}$$

$$\frac{d^2x}{ds^2} = \frac{1}{c^2}\frac{d^2x}{dt^2}, \quad \frac{d^2y}{ds^2} = \frac{1}{c^2}\frac{d^2y}{dt^2}, \quad \frac{d^2z}{ds^2} = \frac{1}{c^2}\frac{d^2z}{dt^2} \tag{33 : 1}$$

Expressed in terms of v's these are (since $v_4 = ct$):

$$\frac{dv_m}{ds} = 0 \; [m = 1, 2, 3] \quad \frac{dv_4}{ds} = 1 \quad \frac{d^2v_m}{ds^2} = \frac{1}{c^2}\frac{d^2v_m}{dt^2} \tag{33 : 2}$$

Let us now give p in (32 : 15) a specific determination —for instance, 1—and proceed to work the equation out. The first thing to do is to observe what values g^{pa}, now become g^{1a}, assumes for $a = 1, 2, 3, 4$. Since space-time is assumed to be Galilean the determinant of the g's is

$$\begin{vmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 \end{vmatrix} = -1 \tag{33 : 3}$$

the terms $g_{11}, g_{12}, g_{13}, g_{14}$ are $-1, 0, 0, 0$, their minors are

$+1, 0, 0, 0$, and the values of g^{11}, g^{12}, g^{13}, g^{14} are, in accordance with the definition in (25 : 3), $-1, 0, 0, 0$. Thus unity is the only value of a which need be taken into account. (In general, $a = p$ is the only value to be taken into account.) Hence (32 : 15) becomes

$$\frac{d^2v_1}{ds^2} - \sum_m \sum_n \tfrac{1}{2}\left(\frac{\partial g_{m1}}{\partial v_n} + \frac{\partial g_{n1}}{\partial v_m} - \frac{\partial g_{mn}}{\partial v_1}\right)\frac{dx_m}{ds}\frac{dx_n}{ds} = 0 \quad (33:4)$$

But from (33 : 2) it appears that dv_m/ds and dv_n/ds are both zero unless $m = 4 = n$, when they are both unity. And since, in (33 : 3), g_{41} is zero, we have finally

$$\frac{d^2v_1}{ds^2} + \frac{1}{2}\frac{\partial g_{44}}{\partial v_1} = 0 \qquad (33:5)$$

or

$$\frac{d^2x}{dt^2} = -\frac{c^2}{2}\frac{\partial g_{44}}{\partial x} \qquad (33:6)$$

with corresponding equations for y and z. Comparing this result with (27 : 1) we see that if we put

$$-\frac{c^2}{2}\frac{\partial g_{44}}{\partial x} = \frac{\partial P}{\partial x} \qquad (33:7)$$

it becomes identical with Newton's law for a permanent gravitational field.

The reader may be puzzled by the fact that although at the beginning of the argument we assumed $g_{44} = 1$, we have ended by expressing the acceleration in terms of its differential coefficient. The explanation is that although g_{44} has approximately the constant value unity, its slight variations are yet the cause of the phenomena of gravitational acceleration. And it is to be noted that in (33 : 6) they are multiplied by the immensely large number c^2.

CHAPTER VIII

THE GRAVITATION POTENTIALS

§ 34. FROM (33 : 7) it appears that the coefficient g_{44} (or rather $c^2g_{44}/2$) plays the part of a potential. For that reason it is convenient, by an extension of the idea, to speak of the whole of the g's as "the gravitation potentials". The solution of any given problem of gravitation requires first that the values of the gravitation potentials shall be ascertained. In this chapter we are to see how that can be done. It should, however, be understood clearly that the method to be illustrated is valid only upon the assumption that the space-time is nearly Galilean, so that the gravitational potentials have very nearly the values set out in (30 : 2, 3, or 4).

§ 35. As a preliminary exercise we will deal with the disc-problem of § 28. Let it be thin and let its mass per unit of surface be M. Then it can easily be proved by the ordinary principles of mechanics (based, of course, on Newton's law of gravitation as well as upon the second law of motion) that the acceleration of a particle situated at distance x from the disc upon its axis is

$$-2\pi GM(1-\cos\alpha) = -2\pi GM\left\{1-\frac{x}{(r^2+x^2)^{\frac{1}{2}}}\right\} \quad (35:1)$$

where r is the radius of the disc, α the angle subtended by it at the point where the particle is situated,

and G the "gravitation constant". Therefore by (33 : 6)

$$\frac{c^2}{2}\frac{\partial g_{44}}{\partial x} = 2\pi\ GM\left\{1 - \frac{x}{(r^2 + x^2)^{\frac{1}{2}}}\right\}$$

whence by integration

$$g_{44} = \frac{4\pi GM}{c^2}\{x - (r^2 + x^2)^{\frac{1}{2}}\} + K$$

To find the constant K we note that, when x is infinite, space-time is Galilean and $g_{44} = 1$. At the same time $x = (r^2 + x^2)^{\frac{1}{2}}$. Thus $K = 1$, and $g_{44} = (1 - P)$ where

$$P = \frac{4\pi GM}{c^2}\{(r^2 + x^2)^{\frac{1}{2}} - x\} \qquad (35:2)$$

To fulfil the condition that g_{44} is nearly unity, P must be a number whose square may be neglected. If the disc is to have the generous extent assigned to it in § 28, this implies that M is small.

Note that the value of g_{44} has been calculated only for points on the axis of the disc. It may, however, be assumed that (35 : 2) holds for all places where the properties of space-time are stratified in the way explained in § 28.

To find the remaining g's it would be tempting to begin by arguing that since there is no acceleration parallel to the disc, B in (28 : 3) must be unity. But a particle on the perpendicular to the disc at a point near its edge would not be "attracted" along that perpendicular; its motion would be directed inwards. It is clear, therefore, that the coefficients of δy^2 and δz^2 must somewhere depart from the Galilean values they possess infinitely far from the disc, though they *may* recover them where the influence of the edge disappears as one approaches

the axis. We must not, therefore, assume that B is unity. To get over the difficulty, we adopt the device of changing our system of linear measurement. Let us take at each place such a linear unit that the former $B\delta y^2$ and $B\delta z^2$ become simply δy^2 and δz^2. Since P is a function of linear measures, it—as well as δx^2—will suffer corresponding changes. No inconsistency with ordinary measurement will, however, be introduced; for when we move to Galilean space-time B returns to the value unity. To put the matter in another way: The ordinary ideas about linear measurement belong to ordinary space-time and must be modified when they are applied to space-time in which the metrical properties are changed. We may use any modification which is convenient, with the restriction that the method adopted must pass continuously into the ordinary method as we approach the Galilean region.

So much for g_{22} and g_{33}. To find g_{11} we argue that if conditions are to remain nearly Galilean, the determinant of the g's must retain its Galilean value at least to the first order of approximation. This will be the case if we take for the new determinant

$$\begin{vmatrix} -(1-P)^{-1} & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & c^2(1-P) \end{vmatrix} = -c^2$$

Thus (28 : 3) becomes

$$\delta s^2 = -(1-P)^{-1}\delta x^2 - \delta y^2 - \delta z^2 + c^2(1-P)\,\delta t^2 \qquad (35:3)$$

For points where x is small in comparison with r we may write

$$P = \frac{4\pi GM}{c^2}(r-x) = \frac{4\pi rGM}{c^2}\left(1-\frac{x}{r}\right)$$

and, since $(1 - P)^{-1} = (1 + P)$ approximately, we have then

$$\delta s^2 = -\left\{1 + \frac{4\pi r GM}{c^2}\left(1 - \frac{x}{r}\right)\right\}\delta x^2 - \delta y^2 - \delta z^2$$
$$+ c^2\left\{1 - \frac{4\pi r GM}{c^2}\left(1 - \frac{x}{r}\right)\right\}\delta t^2 \quad (35:4)$$

§ 36. The foregoing example indicates how we may treat the important problem of determining approximately the gravitation potentials around the sun. In this case we shall work with the polar formula (28 : 4).

By Newton's law of universal gravitation the acceleration of a particle at distance r from the sun's centre is given by

$$\frac{d^2r}{dt^2} = -\frac{GM}{r^2} \qquad (36:1)$$

where M is the whole mass of the sun. Hence by (33 : 6), r being substituted for x,

$$-\frac{c^2}{2}\frac{dg_{44}}{dr} = -\frac{GM}{r^2}$$

and by integration

$$g_{44} = -\frac{2GM}{c^2}\frac{1}{r} + K$$

Since $g_{44} = 1$ when r is infinite, the constant K is unity. Thus

$$g_{44} = \left(1 - \frac{2GM}{c^2}\frac{1}{r}\right) \qquad (36:2)$$
$$= \left(1 - \frac{k}{r}\right)$$

where k is put for $2GM/c^2$.

To find the other g's we argue, as before, that if B is not unity we may take a new r whose values are equal

to the old $r\sqrt{B}$. Also we assume that the determinant of the g's retains its value at least to a first approximation. We thus obtain for the new determinant

$$\begin{vmatrix} -\left(1-\frac{k}{r}\right)^{-1} & 0 & 0 & 0 \\ 0 & -r^2 & 0 & 0 \\ 0 & 0 & -r^2\sin^2\theta & 0 \\ 0 & 0 & 0 & c^2\left(1-\frac{k}{r}\right) \end{vmatrix} = -c^2r^4\sin^2\theta \qquad (36:3)$$

and the required equation is

$$\delta s^2 = -\left(1-\frac{k}{r}\right)^{-1}\delta r^2 - r^2\delta\theta^2 - r^2\sin^2\theta\,\delta\phi^2 + c^2\left(1-\frac{k}{r}\right)\delta t^2 \qquad (36:4)$$

The method by which we have reached (36 : 4) is Einstein's, somewhat differently applied. In view of the great importance of the formula it is desirable to explain his own use of the argument and to show that it leads to the same conclusion.

Einstein sets out to find the g's in a formula which shall express δs^2 in terms of differentials of rectangular co-ordinates whose origin is at the sun's centre. His search for them is guided by the following considerations:

(i) g_{mn} is nearly -1 when $m = n$ and nearly zero when $m \neq n$ $[m, n = 1, 2, 3]$.

(ii) g_{mn} is nearly $+1$ when $m = n = 4$.

(iii) By radial symmetry of the field g_{mn} retains its value when v_m and v_n are both reversed in sign.

(iv) $g_{4n} = 0$ by temporal symmetry unless $n = 4$.

(v) The values -1 and 0 are approached by g_{mn} as r approaches infinity $[m = n = 1, 2, 3]$.

(vi) The determinant of the g's must be nearly -1.

The following array of values is thus indicated:

$$\begin{array}{cccc}
-\left(1+\frac{v_1^2}{r^2}\cdot\frac{k}{r}\right) & -\frac{v_1v_2}{r^2}\cdot\frac{k}{r} & -\frac{v_1v_3}{r^2}\cdot\frac{k}{r} & 0 \\
-\frac{v_2v_1}{r^2}\cdot\frac{k}{r} & -\left(1+\frac{v_2^2}{r^2}\cdot\frac{k}{r}\right) & -\frac{v_2v_3}{r^2}\cdot\frac{k}{r} & 0 \\
-\frac{v_3v_1}{r^2}\cdot\frac{k}{r} & -\frac{v_3v_2}{r^2}\cdot\frac{k}{r} & -\left(1+\frac{v_3^2}{r^2}\cdot\frac{k}{r}\right) & 0 \\
0 & 0 & 0 & +\left(1-\frac{k}{r}\right)
\end{array}$$

where the constant k must be small compared with r.

When the array is regarded as a determinant its value is found to be

$$-\left(1-\frac{k}{r}\right)\left(1+\frac{v_1^2+v_2^2+v_3^2}{r^2}\cdot\frac{k}{r}\right)=-\left(1-\frac{k}{r}\right)\left(1+\frac{k}{r}\right)$$
$$=-1+\frac{k^2}{r^2}$$

i.e. a number which differs only slightly from -1.

These values of the g's (together with δv_1, δv_2 and δv_3) may be converted into polar values by means of the equivalences given in (26 : 2, 3). If the calculation is carried out it will be found that the formula obtained for δs^2 is

$$\delta s^2=-\left(1+\frac{k}{r}\right)\delta r^2-r^2\delta\theta^2-r^2\sin^2\theta\,\delta\phi^2+c^2\left(1-\frac{k}{r}\right)\delta t^2 \qquad (36:5).$$

If k^2/r^2 be neglected this agrees with (36 : 4); for in that case $(1+k/r)=(1-k/r)^{-1}$

§ 37. The foregoing argument assumes that k^2/r^2 is a negligible number. It is important to test this assumption by determining its value.

The earth moves round the sun approximately in a circle whose radius $R = 1{\cdot}49 \times 10^8$ kilometres. If its angular velocity is ω, we have for its acceleration towards the sun

$$\omega^2 R = \frac{GM}{R^2}$$

whence

$$GM = \omega^2 R^3$$

The velocity of light may be taken as 3×10^5 kilometres per second, and ω is determined by the consideration that the earth moves through 2π radians about the sun in 365 days. Thus its angular velocity is found to be $1{\cdot}992 \times 10^{-7}$ rad./sec. With these data we easily obtain

$$k = \frac{2GM}{c^2} = \frac{2\omega^2 R^3}{c^2} = 2{\cdot}94 \qquad (37:1)$$

Since the sun's radius is about 7×10^5 kilometres, k/r is less than $\frac{1}{2}$ of 10^{-5} for all points outside it—a number which may certainly be taken as small.

CHAPTER IX

THE CRUCIAL PHENOMENA

§ 38. We have seen that Einstein's law of motion agrees with Newton's. As a further and much more severe test of the soundness of his principles, Einstein showed that certain phenomena could be deduced from them which could not be accounted for by the older physics. One of these was a feature in the behaviour of the planet Mercury that had long been a puzzle to astronomers. The other two had not hitherto been observed, but ought to be observable; they were the now famous eclipse effect and a certain displacement of the Fraunhofer lines in the solar spectrum. We proceed to show how these "crucial phenomena" could be predicted.

§ 39. *The Spectral Shift.*—The deduction of this phenomenon depends upon the assumption that the vibrations of all atoms of the same element are exactly similar, so that the atoms may be regarded as acting like ideally accurate clocks. Let us consider the beginnings of two consecutive vibrations as two event-particles; then the separation between them, δs, must be the same wherever the atoms may be, and from whatever system it is observed.

Now an atom, though it vibrates, may be regarded as remaining at the same place throughout the course of the

vibration. Thus δr, $\delta\theta$ and $\delta\phi$ are all zero and (36 : 4) reduces to

$$\delta s^2 = c^2 (1 - k/r)\,\delta t^2$$

Since δs remains constant wherever the atom may be situated, this relation proves that the time-length of its vibration is inversely proportional to $(1 - k/r)^{\frac{1}{2}}$. If, then, we consider two similar atoms at distances r_1 and r_2 from the sun's centre, r_2 being the greater, and if their vibration-periods are T_1 and T_2, we have

$$\begin{aligned} T_1/T_2 &= (1 - k/r_2)^{\frac{1}{2}}/(1 - k/r_1)^{\frac{1}{2}} \\ &= 1 + \tfrac{1}{2}k\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \text{ approx.} \end{aligned} \tag{39 : 1}$$

If one of the atoms is in the sun's photosphere and the other in a terrestrial laboratory, $r_1 = 6{\cdot}97 \times 10^5$ kilometres, $r_2 = 1{\cdot}5 \times 10^8$ km.; while from (37 : 1), $k = 2{\cdot}94$. Whence it can be calculated that T_1/T_2 is $1{\cdot}0000021$.

This result means that the atom in the sun's photosphere is vibrating more slowly than the terrestrial atom. According to the theory of the spectrum, the Fraunhofer line of the atom should, therefore, appear rather nearer the red end in the solar spectrum than in the spectrum of the same element in the laboratory. It is claimed that the shift has been verified in the case of cyanogen and magnesium *; but the results do not appear as yet to be generally accepted by physicists.

§ 40. *The Bending of Light.*—The fact that a ray of light ought to deviate from the straight path if it passes near the sun is easily deduced from (36 : 4). From the principle

* Becquerel, *Le principe de relativité*, 1922, p. 241.

of equivalence, applied as in § 31, it follows that the separation between two event-particles along the route of a light-ray is always zero. As regards Galilean space-time, this statement is simply a technical way of expressing (3 : 1), the fundamental relation from which the whole theory of relativity started (cf. § 22); and since the separation between two event-particles is an invariant for all systems and in all circumstances, it holds good also for a gravitational field. Thus when a world line records the history of a pulse of light, (36 : 4) becomes

$$0 = -(1 + k/r)\,\delta r^2 - r^2\delta\theta^2 - r^2 \sin^2\theta\,\delta\phi^2 + c^2(1 - k/r)\,\delta t^2$$

$(1 + k/r)$ being put for $(1 - k/r)^{-1}$ as in (36 : 5); or

$$(1 + k/r)\,\delta r^2 + r^2\delta\theta^2 + r^2 \sin^2\theta\,\delta\phi^2 = c^2(1 - k/r)\,\delta t^2 \quad (40 : 1)$$

In fig. 7 let S be the sun's centre, SAB the initial line

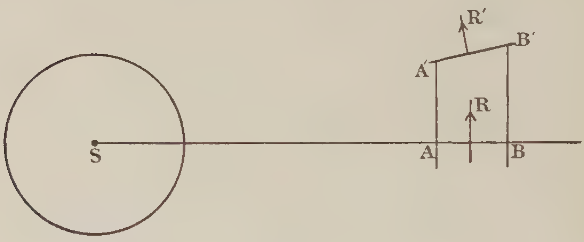

FIG. 7.

from which θ is measured, the page the plane from which ϕ is measured; and let BB' be a ray of light crossing SB at right angles. There is evidently no reason why the ray, as it proceeds, should leave the original plane, so we may put $\delta\phi = 0$ in (40 : 1). Also, since BB' is perpendicular to SB, δr is zero at B. From the standpoint of

an observer at rest with regard to the sun, we then have for the velocity of the ray *,

$$r\frac{d\theta}{dt} = c\,(1 - k/r)^{\frac{1}{2}}$$
$$= c\,(1 - \tfrac{1}{2}k/r) \text{ approx.} \qquad (40:2)$$

from which it appears that the velocity increases as r increases.

Now let AB be a portion of the wave-front crossing SB. Then since the velocity is less at A than at B, AB will swing round into some position $A'B'$ as it moves forward, and will continue to change its direction as it proceeds. Thus a ray, which is normal where it crosses SB at R, will be bent towards the sun at R', and will continue to bend as it pursues its course.

The calculation of the amount of bending can be carried out in the following fairly simple way.†

We must first note that the velocity of light along SB, as observed from a system at rest with regard to the sun, has a value different from that given in (40 : 2). To calculate it we have to put $\delta\theta = 0$ and $\delta\phi = 0$ in (40 : 1), and we then have

$$\delta r/\delta t = c\,(1 - k/r)^{\frac{1}{2}}/(1 + k/r)^{\frac{1}{2}}$$
$$= c\,(1 - k/r) \text{ approx.} \qquad (40:3)$$

* The assumption made here that the velocity of light may vary in amount and direction may seem to contradict the fundamental principle upon which the theory of relativity rests. The contradiction is, however, only apparent. The constancy and rectilinearity of light are established by the Michelson–Morley experiment only for an observer in a Galilean system. Now an observer anywhere on the track of the light-pulse will, by the principle of relativity, judge himself to be in a Galilean region of space-time ; consequently he will judge the speed of the light-pulse to be c and its path to be rectilinear. But this fact does not exclude the possibility that the world line of the light-pulse may be a four-dimensional curve.

† A less elementary but more satisfactory proof is given in § 42.

Before we proceed something must be done to remove this anomaly. The device usually employed is to measure r not from the absolute centre of the sun but from the circumference of a circle of radius $\frac{1}{2}k$ drawn round it. Since the sun's radius is nearly 7×10^5 km. and $\frac{1}{2}k$ is 1·47 km. (37 : 1), no serious error can thus be introduced. Where r appears in (40 : 1) we must now substitute $r + \frac{1}{2}k$; with the result that upon dividing by $(1 + k/r)$ and dropping k^2 wherever it appears the following changes take place:

$$\frac{r^2}{1 + k/r} \text{ becomes } \frac{(r + \frac{1}{2}k)^2}{1 + k/(r + \frac{1}{2}k)}$$

$$= \frac{(r + \frac{1}{2}k)^3}{r + \frac{3}{2}k}$$

$$= r^2\,(r + \tfrac{3}{2}k)\,(r + \tfrac{3}{2}k)^{-1} = r^2$$

while $\dfrac{1 - k/r}{1 + k/r}$ becomes $\dfrac{1 - k/(r + \frac{1}{2}k)}{1 + k/(r + \frac{1}{2}k)}$

$$= \frac{r - \frac{1}{2}k}{r + \frac{3}{2}k} = 1 - \frac{2k}{r}$$

On the other hand $\delta(r + \frac{1}{2}k) = \delta r$; so that δr is unchanged.

With these values substituted, (40 : 1) is transformed into

$$\delta r^2 + r^2\delta\theta^2 + r^2 \sin^2\theta\,\delta\phi^2 = c^2\,(1 - 2k/r)\delta t^2 \qquad (40:4)$$

and the transverse and longitudinal velocities now both become $c\,(1 - 2k/r)^{\frac{1}{2}}$, i.e. approximately $c\,(1 - k/r)$. Equality in the rectangular directions being secured, it follows that the velocity in all other directions is also $c\,(1 - k/r)$.

This result enables us to consider a ray such as PA in fig. 8 as penetrating a series of thin spherical layers, con-

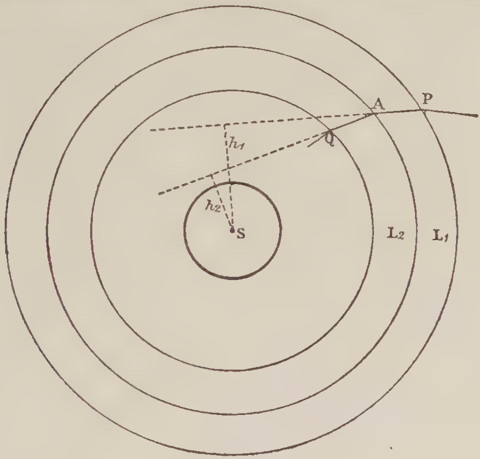

FIG. 8.

centric with the sun, in which its velocity constantly diminishes—in other words, layers whose refractive index constantly increases. Let r_1 and r_2 be the mean radii of two consecutive layers, L_1 and L_2, and let AQ be the ray after refraction at their common surface; also let h_1 and h_2 be the lengths of the perpendiculars upon PA and AQ drawn from S. Then by the theory of light the ratio of the refractive indices of the layers is $(1 - k/r_1)/(1 - k/r_2)$, and by the law of refraction this ratio is equal to h_1/h_2. We have, therefore, the relation

$$h(1 - k/r)^{-1} = h(1 + k/r) = \text{constant} \quad (40:5)$$

all along the path of the ray.

Since a pulse of light, as it moves along its curved path, behaves in a general way like a particle "attracted" by the sun, it is not unreasonable to surmise that its track

may be a parabola or a hyperbola with the sun occupying the focus. In that case the equation of the path would be

$$L/r = 1 + e\cos\theta \qquad (40:6)$$

where the angle θ is measured from the line SA (fig. 9),

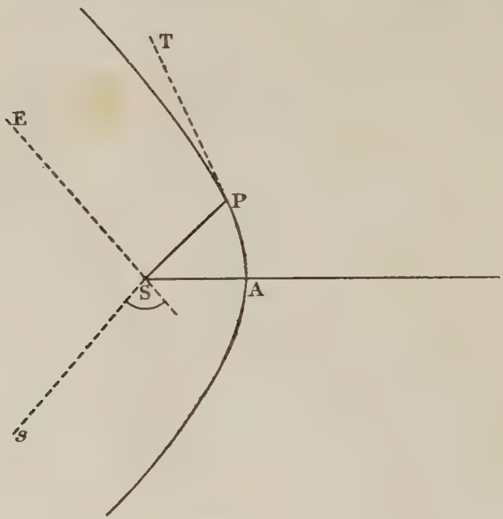

FIG. 9.

e is the eccentricity of the curve and L its semi-latus rectum. If R is the distance SA, then $L = R\,(1 + e)$, since $\cos\theta = 1$ for the point A.

The direction of the ray at a point P is along the tangent PT. Let the angle PSA be α; then the equation of the tangent is

$$\begin{aligned} L/r &= e\cos\theta + \cos(\theta - \alpha) \\ &= (e + \cos\alpha)\cos\theta + \sin\alpha\sin\theta \end{aligned}$$

Hence

$$\begin{aligned} L &= (e + \cos\alpha)\, r\cos\theta + \sin\alpha\, r\sin\theta \\ &= (e + \cos\alpha)\, x + \sin\alpha\, y \qquad (40:7) \end{aligned}$$

From (40 : 7) we deduce by the ordinary formula of co-ordinate geometry that the length h of the perpendicular from the origin S upon the direction of the ray at P is

$$h = \frac{L}{\sqrt{\{(e + \cos a)^2 + \sin^2 a\}}} \qquad (40:8)$$

$$= \frac{L}{\sqrt{(1 + 2e\cos a + e^2)}}$$

Also, at P

$$(1 + k/r) = 1 + \frac{k}{L}\cdot\frac{L}{r}$$

$$= 1 + \frac{k}{L}(1 + e\cos a)$$

$$= \left(1 + \frac{k}{L}\right) + \frac{ek}{L}\cos a$$

putting $\theta = a$ in (40 : 6). Thus (40 : 5) becomes

$$h\left(1 + \frac{k}{r}\right) = \frac{L\left\{\left(1 + \frac{k}{L}\right) + \frac{ek}{L}\cos a\right\}}{\sqrt{(1 + 2e\cos a + e^2)}}$$

$$= \frac{L + k + ek\cos a}{\sqrt{(1 + 2e\cos a + e^2)}}$$

$$= \frac{(R + k) + e(R + k\cos a)}{\sqrt{(1 + 2e\cos a + e^2)}}$$

$$= \text{constant} \qquad (40:9)$$

since $L = R(1 + e)$.

Now when the pulse of light is at A its direction is normal to SA, so that we have $h = r = R$ and

$$h(1 + k/r) = R + k \qquad (40:10)$$

Equating the right-hand sides of (40 : 9, 10) gives

$$(R + k) + e(R + k\cos a)$$
$$= (R + k)\sqrt{(1 + 2e\cos a + e^2)}$$

whence, by a little algebra which may be left to the reader, we arrive at the result

$$e = \frac{R}{k} \cdot \frac{2R + 2k}{2R + k(1 + \cos a)}$$
$$= \frac{R}{k} \text{ very nearly} \qquad (40:11)$$

Now a conic section may be defined as a locus corresponding to (40 : 6) when e is a constant. Since we have just shown that e has very nearly the constant value R/k, we are entitled to deduce that the path of the ray is, at least very nearly, a conic section. And since R/k is greater than unity, it is a hyperbola.

To calculate the amount of bending, we note that from the standpoint of the sun, both the star from which the ray comes and the earth where it arrives may be regarded as at infinity. Thus the angle between the emergent and the arriving ray may be taken as the angle between the asymptotes of the hyperbola.*

* Readers who are familiar with the geometry of the hyperbola may substitute for the above the following less clumsy proof. Let S' be the second focus and h' the length of the perpendicular therefrom upon the tangent at P; also let $S'P = r'$. Then we have

$$r' - r = 2a \qquad hh' = b^2$$

a and b being the lengths of the semi-axes. But the tangent at P bisects the angle SPS'; hence

$$h/r = h'/r'$$

from which it follows, by substitution, that

$$h^2(1 + 2a/r) = b^2, \qquad (40:12)$$

Now from (40 : 10), which is independent of the algebraic argument preceding it, we have, since k/r is small,

$$h^2(1 + 2k/r) = (R + k)^2 \qquad (40:13)$$

From comparison of (40 : 13, 14) we conclude that the path of a light-pulse is (very nearly) a hyperbola whose semi-axes are k and $R + k$. The formula for the angle of bending then follows as in the text.

Now the angle between the y-axis and each asymptote of the hyperbola $x^2/a^2 - y^2/b^2 = 1$ is the angle whose tangent is a/b; but since the angle is extremely small we may identify its tangent with its circular measure. Thus the angle between the two asymptotes is $2a/b$ radians. But $b^2 = a^2(e^2 - 1)$. Hence

$$\begin{aligned} \text{angle of bending} &= 2/\sqrt{(e^2 - 1)} \\ &= 2/\sqrt{(R/k)^2 - 1} \\ &= 2k/R \quad \text{radians very nearly} \end{aligned} \tag{40 : 14}$$

The radius of the sun is 697,000 km. and $k = 2{\cdot}94$ (37 : 1). Substituting these values in (40 : 12), we find that a ray which just grazed the sun's surface would be bent through an angle of $8{\cdot}437 \times 10^{-6}$ rads. or $1''{\cdot}74$; for other rays, the angle, being inversely proportional to R, would be less. Suppose it possible to observe from earth a star whose rays suffer the maximum bending, and which is seen, therefore, as a bead on the sun's rim. Since the star's distance is practically infinite both from the earth and from the sun, the line joining it to the observer may be taken as parallel to one of the asymptotes, while, as we have seen, the light reaches the observer's eye along the other. Thus the observed displacement of the star will be equal to the angle between the asymptotes.

It is popular knowledge that elaborate attempts were made during the total eclipses of 1919 and 1922 to verify the minute displacements of the stars which appeared in photographs during the moment of totality—the measurements being made, of course, by comparison with photographs of the same region of the sky taken some months earlier or later. The technical difficulties were enormous,

but the results obtained in 1919 were nevertheless held to confirm Einstein's prediction in a striking manner. When the measured displacement of each star at distance R' from the sun's centre was multiplied by R'/R so as, in accordance with (40 : 14), to deduce from it the displacement at grazing incidence, the mean value of the bending was found by the Sobral expedition to be $1''\cdot98$ with a probable error of $0''\cdot12$. This means that one-half of the calculated displacements for grazing incidence lay between $2''\cdot1$ and $1''\cdot86$. The corresponding results obtained by the Principe expedition were $1''\cdot61$ with a probable error of $0''\cdot3$—that is, with half the displacements lying between $1''\cdot91$ and $1''\cdot31$.

Now before Einstein had arrived at the conceptions of the general theory of relativity he had already (1911) calculated, on Newtonian principles, that light grazing the sun's edge should be bent through $0''\cdot83$—i.e. about one-half the angle calculated in (40 : 14). The verdict of the 1919 observations is therefore clearly in favour of the validity of Einstein's later theory as against the former*. It has also been reported, since the above was written, that the results of the 1922 expeditions are even more strongly confirmatory of Einstein's prediction.

§ 41. *The Perihelion of Mercury*.—Like all the planets, Mercury moves in an elliptical orbit of which the sun occupies one focus. The point A (fig. 10) at which it is nearest to the sun is called its "perihelion", and the line AB, which is of course the major axis of the ellipse, is

* The reader is, however, reminded that Prof. Whitehead has deduced Einstein's expression for the bending of light-rays without making his assumption that space is modified in the neighbourhood of gravitating matter.

called the "line of apses" or "apsidal line". On the classical theory of gravitation, if Mercury had been the sun's solitary satellite, the apsidal line would have pointed

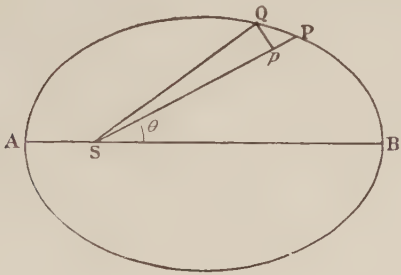

FIG. 10.
$\angle PSQ = \delta\theta$; $pQ = r\delta\theta$; area $PQS = \frac{1}{2}r^2\delta\theta$.

constantly in the same direction among the fixed stars. But there are other planets not so very far away, and since their periods are different from Mercury's they exercise a disturbing influence which results in a slow rotation of the line of apses. The amount of rotation to be accounted for in this way was worked out long ago, but was found to be 43″ per century short of the displacement actually observed. Minute as the discrepancy may appear to the lay mind, it caused the astronomers much perplexity, and many efforts were made to explain it away. The history of the discovery of Neptune suggested that it might be due to an unknown planet circulating within the orbit of Mercury. Optimistic observers even persuaded themselves that they had caught sight of the disturber crossing the sun's disc, and went so far as to christen him Vulcan. But Vulcan refused to confirm their belief in his existence and is now recognized as a mythical

being. One of Einstein's greatest triumphs is to have shown that the unexplained motion of the line of apses may be regarded as an expression of the difference between his own and the classical view of gravitation.

If the distance SP (fig. 10) and the angle PSB are respectively r and θ, and if u is put for $1/r$, then it can be proved * that, on Newton's theory,

$$\left.\begin{aligned} \frac{d^2u}{d\theta^2} + u &= \frac{m}{h^2} \\ r^2\frac{d\theta}{dt} &= h \end{aligned}\right\} \qquad (41:1)$$

where h is (as can be seen from fig. 10) twice the (constant) area swept out by the radius vector in unit time and m represents the GM of § 36. The reader (if the subject-matter is new to him) may easily verify that the differential equation (41 : 1) is satisfied by

$$u = \frac{m}{h^2}(1 + e\cos\theta) \qquad (41:2)$$

from which it appears that the path of the planet is a conic section with semi-latus rectum h^2/m.

From what we saw in § 33 it is to be expected that the formulæ in the theory of relativity which correspond to (41 : 1) will contain s in the place of t. Also Newton's second law, upon which the deduction of (41 : 1) is based, must, of course, be replaced by the law of geodesic motion (32 : 15). Following up these clues, and assuming that the planet's orbit is confined to the $r\theta$-plane, we proceed first to find expressions for $d\theta/ds$ and dt/ds.

(i) For this purpose we take $v_1 = r$, $v_2 = \theta$, $v_3 = \phi$,

* See e.g. Tait & Steele, *Dynamics of a Particle*, ch. v.

$v_4 = ct$. We then deduce from the determinant (36 : 3) the following values :

$$g^{11} = -(1 - k/r),\ g^{22} = -1/r^2,\ g^{33} = -1/(r^2 \sin^2\theta),$$
$$g^{44} = 1/\{c^2(1 - k/r)\} \qquad (41:3)$$

All other values of g^{mn} are zero (see § 33). To find $d\theta/ds$ we put $p = 2$ in (32 : 15) and proceed to calculate the several values of $\{mn, 2\}$, referring for the purpose to (32 : 17). Since $p = 2$, 2 is the only value of a to be taken into account. And on trying in succession the 16 pairs of possible values of m and n, we see that the only values o $\{mn, 2\}$ which are not zero are :

$$\{12, 2\} = 1/r\,; \quad \text{and } \{21, 2\} = 1/r \qquad (41:4)$$

For instance, when $m = 1$ and $n = 2$, (32 : 17) becomes

$$\{12, 2\} = \tfrac{1}{2}g^{22}\left(\frac{\partial g_{12}}{\partial v_2} + \frac{\partial g_{22}}{\partial v_1} - \frac{\partial g_{12}}{\partial v_2}\right)$$

so the only term within the bracket that survives is

$$\partial g_{na}/\partial v_m = \partial g_{22}/\partial r = -d(r^2)/dr = -2r.$$

Hence $$\{12, 2\} = \tfrac{1}{2}\left(-\frac{1}{r^2} \times -2r\right) = \frac{1}{r}$$

and it is evident that $\{21, 2\}$ has the same value. Next let $m = 3$, $n = 4$, so that

$$\{34, 2\} = \tfrac{1}{2}g^{22}\left(\frac{\partial g_{32}}{\partial v_4} + \frac{\partial g_{42}}{\partial v_3} - \frac{\partial g_{34}}{\partial v_2}\right)$$

In this case, since g_{32}, g_{42} and g_{34} are all zero the value of the 3-index symbol is also zero. The other cases can be dealt with in the same way.

Applying these results to the evaluation of (32 : 15) we have

$$\frac{d^2v_2}{ds^2} + \sum_m\sum_n \{mn,\ 2\}\ \frac{dv_m}{ds}\ \frac{dv_n}{ds} = \frac{d^2\theta}{ds^2} + \{12,\ 2\}\ \frac{dr}{ds}\ \frac{d\theta}{ds}$$

$$+ \{21,\ 2\}\ \frac{d\theta}{ds}\ \frac{dr}{ds}$$

$$= \frac{d^2\theta}{ds^2} + \frac{2}{r}\ \frac{dr}{ds}\ \frac{d\theta}{ds}$$

From this point the solution of (32 : 15) proceeds as follows :

$$\frac{d^2\theta}{ds^2} + \frac{2}{r}\ \frac{dr}{ds}\ \frac{d\theta}{ds} = 0$$

$$r^2\ \frac{d^2\theta}{ds^2} + 2r\ \frac{dr}{ds}\ \frac{d\theta}{ds} = 0$$

$$\frac{d}{ds}\left(r^2\ \frac{d\theta}{ds}\right) = 0$$

$$r^2\ \frac{d\theta}{ds} = h \qquad (41 : 5)$$

where h is a constant. Thus we have established the anticipated analogy with the second equation in (41 : 1).

To find dt/ds, we put $p = 4$ in (32 : 15) and have

$$\frac{d^2t}{ds^2} + \sum_m\sum_n \{mn,\ 4\}\ \frac{dv_m}{ds}\ \frac{dv_n}{ds} = 0$$

By the method already illustrated it is easily proved that the only cases in which $\{mn,\ 4\}$ is not zero are those in which $m = 4$ and $n = 1$ or $m = 1$ and $n = 4$. Thus

$$\{14,\ 4\} = \tfrac{1}{2}g^{44}\left(\frac{\partial g_{14}}{\partial v_4} + \frac{\partial g_{44}}{\partial v_1} - \frac{\partial g_{14}}{\partial v_4}\right) = \tfrac{1}{2}g^{44}\ \frac{\partial g_{44}}{\partial v_1}$$

Now $g^{44} = 1/c^2\,(1 - k/r)$ and

$$\frac{\partial g_{44}}{\partial v_1} = c^2\,\frac{d}{dr}\left(1 - \frac{k}{r}\right) = \frac{kc^2}{r^2}$$

Hence $\{14, 4\} = \frac{1}{2}\dfrac{1}{c^2\left(1-\dfrac{k}{r}\right)} \times \dfrac{kc^2}{r^2} = \frac{1}{2}\dfrac{k}{r^2}\left(1-\dfrac{k}{r}\right)^{-1}$

and $\{41, 4\}$ obviously has the same value. In this case, then, (32 : 15) becomes

$$\frac{d^2t}{ds^2} + \{14, 4\}\frac{dr}{ds}\frac{dt}{ds} + \{41, 4\}\frac{dt}{ds}\frac{dr}{ds} = 0$$

that is
$$\frac{d^2t}{ds^2} + \frac{k}{r^2}\left(1-\frac{k}{r}\right)^{-1}\frac{dr}{ds}\frac{dt}{ds} = 0$$

whence
$$\left(1-\frac{k}{r}\right)\frac{d^2t}{ds^2} + \frac{k}{r^2}\frac{dr}{ds}\frac{dt}{ds} = 0$$

$$\frac{d}{ds}\left[\left(1-\frac{k}{r}\right)\frac{dt}{ds}\right] = 0$$

$$\left(1-\frac{k}{r}\right)\frac{dt}{ds} = K \qquad (41:6)$$

To determine the value of the constant K, we note that $dt/ds = K$ when r is infinite. But in that case (36 : 4) reduces to the Galilean form

$$ds^2 = -dr^2 - r^2\delta\theta^2 + c^2\delta t^2 \qquad (41:7)$$

the term in $\delta\phi^2$ being omitted because the motion of the planet is confined to the $r\theta$-plane. Making r and θ constant we deduce from this that $dt/ds = 1/c$; whence $K = 1/c$. Thus (41 : 6) becomes

$$\frac{dt}{ds} = \frac{1}{c\left(1-\dfrac{k}{r}\right)} \qquad (41:8)$$

We now divide the equation

$$\delta s^2 = -\left(1-\frac{k}{r}\right)^{-1}\delta r^2 - r^2\delta\theta^2 + c^2\left(1-\frac{k}{r}\right)dt^2$$

throughout by δs^2, and substituting from (41 : 5, 8) obtain

$$1 = -\left(1 - \frac{k}{r}\right)^{-1}\left(\frac{dr}{ds}\right)^2 - r^2\left(\frac{h}{r^2}\right)^2 + c^2\left(1 - \frac{k}{r}\right)\cdot\frac{1}{c^2\left(1 - \frac{k}{r}\right)^2}$$

i.e. $$\left(\frac{dr}{ds}\right)^2 + \left(1 - \frac{k}{r}\right)\frac{h^2}{r^2} - \frac{k}{r} = 0 \qquad (41:9)$$

We now make two substitutions in succession :

(i) $\dfrac{dr}{ds} = \dfrac{dr}{d\theta}\cdot\dfrac{d\theta}{ds} = \dfrac{h}{r^2}\cdot\dfrac{dr}{d\theta}$;

(ii) $u = \dfrac{1}{r}$, whence $\dfrac{du}{d\theta} = -\dfrac{1}{r^2}\dfrac{dr}{d\theta}$ and $\dfrac{h}{r^2}\dfrac{dr}{d\theta} = -h\dfrac{du}{d\theta}$

With these changes (41 : 9) becomes

$$h^2\left(\frac{du}{d\theta}\right)^2 + (1 - ku)\,h^2u^2 - ku = 0$$

Differentiated with regard to θ, this equation yields

$$2h^2\left(\frac{du}{d\theta}\right)\frac{d^2u}{d\theta^2} + 2h^2u\frac{du}{d\theta} - 3kh^2u^2\frac{du}{d\theta} - k\frac{du}{d\theta} = 0$$

or $$\frac{d^2u}{d\theta^2} + u = \frac{k}{2h^2} + \frac{3}{2}ku^2 \qquad (41:10)$$

Comparison with (41 : 1) shows that the term $\frac{3}{2}ku^2$ is the one not foreseen by the Newtonian theory. The agreement of the term $k/2h^2$ with the m/h^2 of (41 : 1) is established when one observes that in the former $h = r^2\dfrac{d\theta}{ds}$ and in the latter $h = r^2\dfrac{d\theta}{dt}$. Since $\delta s = c\delta t$ *, the h of

* That is for a body moving so slowly as Mercury moves in comparison with the speed of light. Cf. § 33.

(41 : 1) is c times the h of (41 : 10). Again, in (41 : 1) $m = GM$, while in (41 : 10) $k = 2GM/c^2$. Thus, if we dash the h of (41 : 10) to distinguish it from the h of (41 : 1), we have

$$\frac{k}{2h'^2} = \tfrac{1}{2}\left(\frac{2GM}{c^2}\right)\frac{c^2}{h^2}$$

$$= \frac{m}{h^2}$$

Now the ratio of the second to the first term on the right of (41 : 10) is $3u^2h^2$. To estimate its value we have $r = 1/u = 5{\cdot}79 \times 10^7$ km., while by (41 : 5) and the footnote on p. 120, $h = (r^2/c)\,(d\theta/dt)$. Since Mercury revolves through 2π radians about the sun in 88 days the mean value of $d\theta/dt$ is $2\pi/(88 \times 24 \times 60 \times 60)$. Also c is 3×10^5. From these data $h = 9{\cdot}2 \times 10^3$ and $3u^2h^2$ is about $7{\cdot}6 \times 10^{-8}$. Thus Einstein's correction of the Newtonian formula is very small. It follows that a first approximation to the solution of (41 : 10) may be obtained by ignoring the term $\frac{3}{2}ku^2$ and putting

$$u_1 = \frac{k}{2h^2}(1 + e\cos\theta) \qquad (41 : 11)\ *$$

which will be found upon differentiation to satisfy

$$\frac{d^2u_1}{d\theta^2} + u_1 = \frac{k}{2h^2}$$

To obtain a more accurate solution, we assume that $u = u_1 + u_2$, and take the equation

$$\frac{d^2u_2}{d\theta^2} + u_2 = \tfrac{3}{2}ku^2$$

* Strictly speaking, the more general cos $(\theta - \alpha)$ should be used instead of cos θ, but the simpler expression will suffice for our purpose.

which, when the approximate value of u (*i.e.* u_1) is substituted on the right, becomes

$$\frac{d^2u_2}{d\theta^2} + u_2 = \tfrac{3}{2}k\left\{\frac{k^2}{4h^4}(1 + e\cos\theta)^2\right\}$$

$$= \frac{3k^3}{8h^4}(1 + \tfrac{1}{2}e^2) + \frac{3k^3e}{4h^4}\cos\theta + \frac{3k^3e^2}{16h^4}\cos 2\theta$$

$$= \quad A \quad + \quad B\cos\theta \quad + \quad C\cos 2\theta$$

We may now put $u_2 = w_1 + w_2 + w_3$, where each of the three components of u_2 corresponds to one of the three terms on the right. The first gives an equation which would add to the approximate value of u a correction of the same form as (41 : 11), but so small as to be negligible. As regards the third term, we note that $w_3 = -\frac{1}{3}C\cos 2\theta$ is a solution of

$$\frac{d^2w_3}{d\theta^2} + w_3 = C\cos 2\theta$$

But inasmuch as it passes through its periodic series of values twice as quickly as the right-hand side of (41 : 11), any disturbance which it might produce during one-half of the planet's revolution about the sun would be wiped out during the second half. We are left, therefore, with the term $B\cos\theta$. Now it is easily seen that $\frac{1}{2}B\theta\sin\theta$ is a solution of

$$\frac{d^2w_2}{d\theta^2} + w_2 = B\cos\theta$$

and we have here a term which does not simply pass through a periodic series of deviations from a constant mean value, but has an *increasing* mean value —for if θ be taken as zero at a given epoch, it will increase by 2π in every revolution of the planet. However small B

may be, the term $\frac{1}{2}B\theta\sin\theta$ will therefore produce in time a visible effect upon the orbit. Adding this term to u_1 we have

$$u = \frac{k}{2h^2}\left\{1 + e\left(\cos\theta + \frac{3}{4}\frac{k^2}{h^2}\,\theta\sin\theta\right)\right\}$$
$$= \frac{k}{2h^2}\{1 + e\sec\alpha\cos(\theta - \alpha)\}$$
$$(41:12)$$

where $\tan\alpha = 3k^2\theta/4h^2$. Now the meaning of (41 : 12) is that the planet is at its perihelion distance $k/2h^2$ when $\theta = \pi + \alpha$ instead of when $\theta = \pi$, as is implied by (41 : 11). And since α increases with the time this shifting of the line of apses will be continuous.

To determine the rate at which the apsidal line revolves we note that by § 37, $k = 2\cdot94$ and that, as calculated above, $h = 9\cdot2 \times 10^3$. With these data $\frac{3}{4}k^2/h^2 = 7\cdot66 \times 10^{-8}$. We can now find how much the perihelion advances during one revolution by putting $\theta = 360°$.

Thus $$\frac{3k^2}{4h^2}\theta = 0''\cdot099$$

In a century Mercury accomplishes $(100 \times 365)/88$ revolutions; whence it appears that the centennial advance of the perihelion is about $41''\cdot2$. A more accurate calculation yields the value $42''\cdot9$, which is precisely what the astronomers require.

§ 42. Following Professor G. B. Jeffery * we can now deduce the existence and amount of the bending of light-rays by the sun by a method which avoids the device we were compelled to use in § 40 in order to make the velocity of light uniform in all directions.

* *Phil. Mag.*, September 1920.

We may regard a ray as the path traced by a "particle" of light in conformity with (41 : 5, 10). Since in this case δs is zero (see p. 106), (41 : 5) shows that h is infinitely great. Consequently the term k/h^2 disappears from (41 : 10), and the equation reduces to

$$\frac{d^2u}{d\theta^2} + u = \tfrac{3}{2}ku^2 \qquad (42:1)$$

In view of the magnitude of the distances involved, we may, as before, treat $\frac{3}{2}ku^2$ as a small quantity. The first approximation to u will therefore be obtained by putting $u = u_1 + u_2$ and solving the equation

$$\frac{d^2u_1}{d\theta^2} + u_1 = 0 \qquad (42:2)$$

The simplest solution admissible is $u_1 = \cos\theta/R$, where R is the value of r $(= 1/u)$ when $\theta = 0$, and is therefore the distance SA in fig. 9. Substituting in (42 : 1) we obtain

$$\frac{d^2u_2}{d\theta^2} + u_2 = \frac{3k}{2R^2}\cos^2\theta \qquad (42:3)$$

as the equation from which the correction is to be calculated. The expression

$$\frac{k^2}{R^2}(1 - \tfrac{1}{2}\cos^2\theta)$$

is a solution of this equation (as the reader may verify by differentiation), and we shall adopt it as the required correction. Thus the more complete solution of (42 : 1) becomes

$$u = \frac{k}{R^2} + \frac{\cos\theta}{R} - \frac{k^2}{2R^2}\cos^2\theta \qquad (42:4)$$

Now, when r is infinite (and u zero), the values of θ which satisfy (42 : 4) will be the directions Ss and SE through S in fig. 9, parallel to the directions of the ray at infinity—i.e. to the directions in which it leaves the star and arrives at the earth. Putting $u = 0$, we have the quadratic equation

$$\frac{k}{2R}\cos^2\theta - \cos\theta - \frac{k}{R} = 0 \qquad (42:5)$$

from which we obtain

$$\cos\theta = \frac{1 \pm \sqrt{\left(1 + \frac{2k^2}{R^2}\right)}}{k/R} = \{1 \pm (1 + k^2/R^2)\}\cdot\frac{R}{k} = -\,k/R$$

the root corresponding to the *plus* sign being greater than unity and therefore inadmissible.

Since k/R is small, we may put $\sin(k/R) = k/R$. If we then put

$$\begin{aligned}\cos\theta &= \cos\left[\pm\left(\tfrac{1}{2}\pi + k/R\right)\right]\\ &= -\sin(k/R)\\ &= -\,k/R\end{aligned}$$

it appears that the two values of θ which satisfy (42 : 3) are $\pm\left(\frac{\pi}{2} + k/R\right)$. Thus the total bending of the ray proves to be $2k/R$ as in (40 : 14).

CHAPTER X

THE TENSOR METHOD

§ 43. The argument of the last three chapters may be summarized as follows. In Chapter VII we sought a law of motion that should be universally valid: that is, valid for all systems of reference, whatever their relative motion, and for both types of space-time, Galilean and non-Galilean. This law we found in the principle that the world line of a free particle is always a geodesic. Before passing on we tested the soundness of the reasoning which led us to that principle by proving that Newton's second law of motion may be regarded as an approximate expression of it, true for particles moving with relatively small speed in the system of reference and in regions of space-time little different from the homaloidal or Galilean.

Our faith in the geodesic law thus confirmed, we proceeded in Chapter VIII to deduce from it a formula expressing the modifications of space-time to be expected within a finite distance of a solitary mass such as the sun. The highly important formula (36 : 4) is the one we adopted.

Lastly, we saw in Chapter IX how these modifications of space-time imply certain phenomena which the older dynamics and physics could not foresee, and learnt that in two, if not all of the three, cases observation has confirmed Einstein's predictions and thus justified his main principles.

Now although the results of the inquiry we undertook under Einstein's guidance appear to be as true as they are striking, the method by which we reached them is not entirely satisfactory. The weak point lies in the second stage where it was assumed, in deducing (36 : 4), that the uniformity of space-time is disturbed only slightly by the presence of matter. A formula whose validity is limited by such an assumption cannot claim to inherit the dignity hitherto granted to Newton's law of gravitation. Before our task can be regarded as completed we must, therefore, find some means either of proving that (36 : 4) is valid in all circumstances or else of correcting its deficiencies in order to make it so.

The nature of the problem may be briefly stated. We assume the results of the restricted theory of relativity and the principle of equivalence which assures us that those results may be applied to any sufficiently limited region of space-time. In particular we assume that the separation, δs, between two given near event-particles is an invariant for all systems, and that it may always be expressed by a quadratic formula whose most general form is

$$\delta s^2 = \sum_m \sum_n g_{mn} \delta v_m \delta v_n$$

We assume, further, that the gravitational properties of any given region of space-time are expressed in the g's, i.e. either by their constancy or by their form if, as is in general the case, they are functions of the co-ordinates.

All these things assumed, our task is to find criteria by means of which, given the distribution of masses in space-time, the g's can be determined not by plausible guessing but by a rigid deductive process whose results

will have full universality. The formula for geodesic motion (32 : 15) is a criterion of this kind; but we have already found that, by itself, it is insufficient to determine the g's. We need, then, others; and, in the light of the preceding inquiry, it is pretty clear where we should look for them. If they are to be found at all, they should emerge from a minuter scrutiny of the conditions under which formulæ which express spatio-temporal properties may claim validity in all systems of reference. As the result of such a scrutiny we shall find that the desired criteria may take the form of equations between the g's of the kind called by mathematicians "tensor equations". In the present and following chapters we proceed to elucidate this statement, and, first, to explain the nature and exhibit the relevant properties of tensors.*

§ 44. *First Order Tensors.*—In Chapter VI, § 24 (i), it was, in effect, shown that if the separation or interval between two event-particles is δs, its components in the V-system of coordinates are connected with its components in any other U-system by the four equations of transformation typified in (23 : 2):

$$\delta v_m = \sum_a \delta u_a \frac{\partial v_m}{\partial u_a}$$

Again, it was shown in § 24 (ii) that, if P is of the nature of a potential, its gradients parallel to the axes in the V-system are connected with the gradients parallel to the axes in the U-system by the equations (23 : 4):

$$\frac{\partial P}{\partial v_m} = \sum_a \frac{\partial P}{\partial u_a} \frac{\partial u_a}{\partial v_m}$$

* It should be noted that the use of the term tensor in the theory of relativity is not the same as its use in connexion with quaternions.

There is no reason to suppose that these relations are confined to the instances given in § 24. One may conceive in a perfectly general manner a character correlated with point-instants in such a way (*a*) that it has four components associated with the four axial directions in any system, and (*b*) that the law of transformation of the components from the U-system to the V-system is expressed by four equations of the same pattern as (23 : 4). If we write ${}_uT_1$, ${}_uT_2$, etc., for the components in the U-system and ${}_vT_1$, ${}_vT_2$, etc., for the components of the same character in the V-system, then we should have for the law of transformation

$$ {}_vT_m = \sum_a {}_uT_a \frac{\partial u_a}{\partial v_m} \qquad (44:1)$$

Similarly we may conceive in general a character whose components in the V-system, which we will write ${}_vT^1$, ${}_vT^2$, etc., are connected with those of the U-system by the relation exhibited in (23 : 2), viz.:

$$ {}_vT^m = \sum_a {}_uT^a \frac{\partial v_m}{\partial u_a} \qquad (44:2)$$

In both these cases the character would be called a " tensor-character of the first order ". By that statement is meant: (i) that the character has four components, one related to each of the four axial directions in any system, and (ii) that a given component in any given system is a linear function of the four components in any other given system, the function having either the form exhibited in (44 : 1) or that shown in (44 : 2).

It will be observed that there is an important difference between the forms of the two linear functions—namely, that the partial differential coefficients in (44 : 2) are those

of (44 : 1) inverted. To mark this difference we speak of the components involved in (44 : 1) as elements of a "covariant" first order tensor; and of those in (44 : 2) as elements of a "contravariant" first order tensor. This difference is indicated in the symbols by the position of the suffixes which show the axis to which the component is related. The system to which the components belong is indicated by a prefix, which may be omitted when only one system is in question. To avoid prolixity we shall use the phrase "the tensor ${}_vT_m$" instead of "the tensor in the V-system whose component in the m-direction is T_m".

§ 45. *Tensors of Higher Orders.*—Let ${}_vA_m$ and ${}_vB_n$ be components of two covariant tensors of the first order, so that

$$ {}_vA_m = {}_uA_1 \frac{\partial u_1}{\partial v_m} + {}_uA_2 \frac{\partial u_2}{\partial v_m} + {}_uA_3 \frac{\partial u_3}{\partial v_m} + {}_uA_4 \frac{\partial u_4}{\partial v_m} $$

and $$ {}_vB_n = {}_uB_1 \frac{\partial u_1}{\partial v_n} + {}_uB_2 \frac{\partial u_2}{\partial v_n} + {}_uB_3 \frac{\partial u_3}{\partial v_n} + {}_uB_4 \frac{\partial u_4}{\partial v_n} $$

then the product of the two left-hand components is expressible as the sum of 16 terms of the form ${}_uA_a {}_aB_b \frac{\partial u_a}{\partial v_m} \frac{\partial u_b}{\partial v_n}$. Let this product be represented by the symbol ${}_vT_{mn}$; then we have

$$ {}_vT_{mn} = \sum_a \sum_b {}_uT_{ab} \frac{\partial u_a}{\partial v_m} \frac{\partial u_b}{\partial v_n} \qquad [m, n, a, b = 1, 2, 3, 4] \tag{45 : 1} $$

and the 16 values of ${}_vT_{mn}$ are said to be components of a covariant tensor of the second order. It will be noted that four of these components are connected with each of the four coordinate axes.

Similarly, the product of two contravariant tensors of the first order yields a contravariant tensor of the second order whose formula of transformation is

$$_vT^{mn} = \sum_a \sum_b {_uT^{ab}} \frac{\partial v_m}{\partial u_a} \frac{\partial v_n}{\partial u_b} \qquad (45:2)$$

In the same way we can proceed to tensors, covariant or contravariant, of the third or any higher order n, the number of components being 4^n. Also we may, by analogy, recognize the existence of tensors of zero order. Since $4^\circ = 1$, a tensor of zero order will consist of a single "component" in each system, and that component must be related indifferently to all four coordinate axes. In other words, it must be a scalar (cf. § 23 (ii)) or non-directed number, such as a temperature or a potential. Moreover, since the transformation formula now degenerates into

$$_vT = {_uT}$$

it is an invariant—that is, a character which has the same measure in all systems. A tensor of zero order may be counted as either covariant or contravariant.

Lastly, we have "mixed tensors" produced by multiplying covariant and contravariant tensors together. For instance, the product of $_vA_{mnp}$, which is one of the 4^3 components of a third order covariant tensor, by $_vB^q$, which is a component of a first order contravariant tensor, yields the transformation-formula

$$_vA_{mnp}B^q = \sum_a \sum_b \sum_c \sum_d {_uA_{abc}B^d} \frac{\partial u_a}{\partial v_m} \frac{\partial u_b}{\partial v_n} \frac{\partial u_c}{\partial v_p} \frac{\partial v_q}{\partial u_d} \quad (45:3)$$

Thus the product of these two tensors yields a tensor of the fourth order, partly covariant and partly contravariant, whose components may be symbolized as $_vT^q_{mnp}$.

§ 46. If these transformation-formulæ are to be valid for the theory of relativity they must be reversible. They can easily be shown to pass this test. For instance, let

$$ {}_vT_m = \sum_a {}_uT_a \frac{\partial u_a}{\partial v_m} $$

then

$$ \sum_m {}_vT_m \frac{\partial v_m}{\partial u_a} = \sum_m \left[\sum_a {}_uT_a \frac{\partial u_a}{\partial v_m} \right] \frac{\partial v_m}{\partial u_a} $$

$$ = \sum_a \left[{}_uT_a \sum_m \frac{\partial u_a}{\partial v_m} \frac{\partial v_m}{\partial u_a} \right] $$

But by (23 : 7) $\sum_m (\partial u_a/\partial v_m)\ (\partial v_m/\partial u_a) = 1$. Hence if we assign a definite value (1, 2, 3 or 4) to a, the right-hand side of the equation reduces to ${}_uT_a$, and we have

$$ {}_uT_a = \sum_m {}_vT_m \frac{\partial v_m}{\partial u_a} $$

A similar proof can be applied to tensors of any type and order.

§ 47. Again, the theory of relativity requires that if a certain formula governs the transformation of a tensor from the U-system to (say) both the V-system and the W-system, then the same formula shall govern transformation from the V-system to the W-system. This condition is also satisfied by the tensor-law.

For example, let

$$ {}_vT^m = \sum_a {}_uT^a \frac{\partial v_m}{\partial u_a} \text{ and } T^n = \sum_a {}_uT^a \frac{\partial w_n}{\partial u_a} $$

then by § 46 we have

$$ {}_uT^a = \sum_m {}_vT^m \frac{\partial u_a}{\partial v_m} $$

and
$$
\begin{aligned}
{}_wT^n &= \sum_a \left[\sum_m {}_vT^m \frac{\partial u_a}{\partial v_m} \right] \frac{\partial w_n}{\partial u_a} \\
&= \sum_m \left[{}_vT^m \sum_a \frac{\partial w_n}{\partial u_a} \frac{\partial u_a}{\partial v_m} \right] \\
&= \sum_m {}_vT^m \frac{\partial w_n}{\partial v_m} \text{ by } (23:4)
\end{aligned}
$$

A similar proof can be applied to tensors of any type and order.

§ 48. Let ${}_vA_m$ and ${}_vB_m$ be corresponding components of two covariant tensors of the first order ; then

$$
\begin{aligned}
{}_v(A_m \pm B_m) &= \sum_a {}_uA_a \frac{\partial u_a}{\partial v_m} \pm \sum_a {}_uB_a \frac{\partial u_a}{\partial v_m} \\
&= \sum_a {}_u(A_a \pm B_a) \frac{\partial u_a}{\partial v_m}
\end{aligned}
$$

Hence ${}_v(A_m \pm B_m)$ may be regarded as the m-component of a tensor which is the sum (or difference) of the original tensors. This result may evidently be extended to tensors of any type and order.

§ 49. Let each of the components of a tensor in a given system be zero. Then it follows from the law of transformation that each of the components in any other system will also be zero. This observation, though so simple, is of the highest importance ; for, as we shall see later, the whole value of the tensor-method in the theory of relativity depends upon it.

§ 50. In § 45 we exhibited tensors of higher order as the products of tensors of lower order. It must not, however, be assumed that a tensor of higher order is necessarily the product of other tensors. The sufficient mark of a tensor of any order is its conformity with the

law either of covariant or of contravariant transformation* from one system to another.

Nevertheless, it can be shown that a tensor of the second or higher order can always be expressed as the *sum* of a product of tensors of lower order. Consider a covariant second order tensor in any system, and let its components be set out in the following array:

$$\begin{array}{cccc} T_{11} & T_{21} & T_{31} & T_{41} \\ T_{12} & T_{22} & T_{32} & T_{42} \\ T_{13} & T_{23} & T_{33} & T_{43} \\ T_{14} & T_{24} & T_{34} & T_{44} \end{array}$$

Now choose a first order tensor whose components A_1, A_2, A_3, A_4 are respectively T_{11}, T_{21}, T_{31}, T_{41}, and another whose component $A'_1 = 1$, while its remaining components are all zero. Then the products $A_m A'_n$ will yield 16 terms, of which the four corresponding to $n = 1$ will be the top row in the above array, while all the rest will be zero. Again select a tensor whose components B_m are respectively T_{12}, T_{22}, T_{32}, T_{42}, and with it another tensor whose component B'_2 is unity while the remaining components are all zero. Then the products $B_m B'_n$ will account for the second row of the array. Following the same principle, choose tensors whose products will account similarly for the third and fourth rows. Then it is clear that, for all values of m and n,

$$T_{mn} = A_m A'_n + B_m B'_n + C_m C'_n + D_m D'_n \tag{50 : 1}$$

* In the case of a mixed tensor, with both laws.

CHAPTER XI

RESTRICTION (OR CONTRACTION) OF TENSORS

§ 51. CONSIDER again the product (45 : 3) of the covariant third order tensor A_{mnp} by the first order contravariant tensor B^q. Among the 4^4 components of the resulting mixed fourth order tensor there will be 16 in which $p = q = 1$, 16 more in which $p = q = 2$, and so on. For any one of these groups we shall have

$$
\begin{aligned}
{}_vT^p_{mnp} &= {}_v(A_{mnp} \times B^p) = \sum_a \sum_b \sum_c {}_u(A_{abc} \times B^c) \frac{\partial u_a}{\partial v_m} \frac{\partial u_b}{\partial v_n} \frac{\partial u_c}{\partial v_p} \frac{\partial v_p}{\partial u_c} \\
&= \sum_a \sum_b \sum_c {}_uT^c_{abc} \frac{\partial u_a}{\partial v_m} \frac{\partial u_b}{\partial v_n} \frac{\partial u_c}{\partial v_p} \frac{\partial v_p}{\partial u_c}
\end{aligned}
$$

Now select from the four groups the terms in which m and n have a particular pair of values (e.g. $m = 2$, $n = 3$), and add them together. There will, of course, be four of them, one in each group, and their sum will be

$$
\begin{aligned}
\sum_p {}_vT^p_{mnp} &= \sum_a \sum_b \sum_c {}_uT^c_{abc} \frac{\partial u_a}{\partial v_m} \frac{\partial u_b}{\partial v_n} \sum_p \frac{\partial u_c}{\partial v_p} \frac{\partial v_p}{\partial u_c} \\
&= \sum_a \sum_b \left[\sum_c {}_uT^c_{abc} \right] \frac{\partial u_a}{\partial v_m} \frac{\partial u_b}{\partial v_n} \text{ by } (23:7)
\end{aligned}
$$

Thus it appears that $\sum_p {}_vT^p_{mnp}$ is the mn-component of a covariant tensor of the second order. Since it consists of terms of ${}_vA_{mnp} \times {}_vB^q$ in which $p = q$, it is called a " restricted * product " of the given tensors.

* This is Professoı Whitehead's term. Einstein, adopting a term of Grassmann's, calls it the "inner product".

In conformity with the first paragraph of § 50, a mixed tensor, whether or not the product of other tensors, will be said to be restricted when we pick out and add together in sets, as in the preceding example, the components in which a covariant and a contravariant index have the same values. Each sum, characterized by a particular set of values of the remaining indices, constitutes one component of the restricted tensor.

What is here called "restriction" is called by Professor Eddington "contraction". Einstein names the process "Verjüngung", that is, "rejuvenescence".

It will be seen that if the components of a mixed tensor have r covariant and s contravariant indices, the order of the tensor derived from it by restriction (contraction) is $(r-1)+(s-1)=r+s-2$. The values of $(r-1)$ and $(s-1)$ determine whether it is covariant, contravariant or still mixed.

Consider next the product (fifth order) of the tensors A_{mnp} and B^{qr}. Among the 4^5 terms there will be 4^3 in which $n=q$ and $p=r$ at the same time. These may be divided into four groups of 16, all the members of a given group having the same value for m. By an extension of the preceding argument, the sum of the terms in any one group will be the m-component of a first order tensor whose law of transformation is

$$\sum_n\sum_p {}_vT^{np}_{mnp} = \sum_a\left[\sum_b\sum_c {}_uT^{bc}_{abc}\right]\frac{\partial u_a}{\partial v_m}$$

The product of the original tensors is in this case said to have been doubly restricted. The process is evidently capable of further repetition with tensors of sufficiently high order.

§ 52. Of special interest is the case when the numbers

of the covariant and the contravariant indices are the same, and we pick out the terms in which each covariant index is equal to its corresponding contravariant index. Consider, for instance, the product of the tensors A_{mn} and B^{pq}. Here the doubly restricted product is a tensor whose components are transformed in accordance with the law

$$\sum_m \sum_n {}_vA_{mn}B^{mn} = \sum_a \sum_b {}_uA_{ab}B^{ab}$$

That is, it is an invariant scalar.

In connexion with this result one naturally remembers the invariant

$$\delta s^2 = \sum_m \sum_n g_{mn}\, \delta v_m \delta v_n$$

As we saw in § 44, δv_m is a first order contravariant tensor; so the product $\delta v_m \delta v_n$ must be a second order contravariant tensor. If the preceding theorem is true conversely as well as directly, it would follow that the 16 values of g_{mn} are the components of a covariant tensor of the second order. In that case, when the g's are given for one co-ordinate system, say the U-system, their values in another system, say the V-system, would be given by the relation

$${}_vg_{mn} = \sum_a \sum_b {}_ug_{ab} \frac{\partial u_a}{\partial v_m} \frac{\partial u_b}{\partial v_n}$$

Now by reason of the invariance of δs^2 we have

$$\sum_a \sum_b {}_ug_{ab}\delta u_a \delta u_b = \delta s^2 = \sum_m \sum_n {}_vg_{mn}\delta v_m \delta v_n \qquad (52:1)$$

and since $\delta u_a \delta u_b$ is a contravariant second order tensor

$$\delta u_a \delta u_b = \sum_m \sum_n \delta v_m \delta v_n \frac{\partial u_a}{\partial v_m} \frac{\partial u_b}{\partial v_n} \qquad (52:2)$$

Hence

$$\sum_a\sum_b {}_u g_{ab}\delta u_a \delta u_b = \sum_a\sum_b {}_u g_{ab} \sum_m\sum_n \delta v_m \delta v_n \frac{\partial u_a}{\partial v_m}\frac{\partial u_b}{\partial v_n}$$

$$= \sum_m\sum_n \left[\sum_a\sum_b {}_u g_{ab} \frac{\partial u_a}{\partial v_m}\frac{\partial u_b}{\partial v_n}\right] \delta v_m \, \delta v_n \qquad (52:3)$$

Comparing (52 : 3) with (52 : 1) we have

$$\sum_m\sum_n [{}_v g_{mn}]\, \delta v_m \delta v_n = \sum_m\sum_n \left[\sum_a\sum_b {}_u g_{ab} \frac{\partial u_a}{\partial v_m}\frac{\partial u_b}{\partial v_n}\right] \delta v_m \delta v_n$$

But this equality holds good for all possible values of δv_m and δv_n. We may conclude, therefore, that

$$ {}_v g_{mn} = \sum_a\sum_b {}_u g_{ab} \frac{\partial u_a}{\partial v_m}\frac{\partial u_b}{\partial v_n} \qquad (52:4)$$

that is, that g_{mn} is, as we surmised, a component of a second order covariant tensor.

§ 53. It was shown in § 25 (v) that if m has a fixed value (1, 2, 3, or 4)

$$\sum_n g_{mn} g^{mn} = 1$$

whence it follows that

$$\sum_m\sum_n g_{mn} g^{mn} = 4$$

This result, being simply an algebraic identity, holds good in all coordinate systems. Hence, by an obvious modification of the argument of § 52, it can be shown that since g_{mn} is a covariant tensor of the second order, g^{mn} must be a contravariant tensor of the same order. Thus its law of transformation is found to be

$$ {}_v g^{mn} = \sum_a\sum_b {}_u g^{ab} \frac{\partial v_m}{\partial u_a}\frac{\partial v_n}{\partial u_b} \qquad (53:1)$$

§ 54. The theorems of the two preceding articles are special cases of a more general truth which may be illustrated by the following example. Let the equality

$$T_{mnp} = A_{mn} T_p$$

hold good for every coordinate system, T_{mnp} and T_p being known to be components of covariant tensors of which the second is completely arbitrary. Then it can be proved that A_{mn} is a component of a second order covariant tensor.

Since the product is a third order covariant tensor

$$_vA_{mn}T_p = \sum_a \sum_b \sum_c {}_uA_{ab}T_c \frac{\partial u_a}{\partial v_m} \frac{\partial u_b}{\partial v_n} \frac{\partial u_c}{\partial v_p}$$

Also

$$_uT_c = \sum_q {}_vT_q \frac{\partial v_q}{\partial u_c} \qquad [q = 1, 2, 3, 4]$$

Hence

$$_vA_{mn}T_p = \sum_q {}_vT_q \sum_c \frac{\partial v_q}{\partial u_c} \frac{\partial u_c}{\partial v_p} \sum_a \sum_b {}_uA_{ab} \frac{\partial u_a}{\partial v_m} \frac{\partial u_b}{\partial v_n}$$

But by (23 : 7) $\sum_c \frac{\partial v_q}{\partial u_c} \frac{\partial u_c}{\partial v_p} = 0$ unless q has the (fixed) value p, when its value is unity. In that case $\sum_q {}_vT_q = {}_vT_p$; hence

$$_vA_{mn}T_p = \left[\sum_a \sum_b {}_uA_{ab} \frac{\partial u_a}{\partial v_m} \frac{\partial u_b}{\partial v_n} \right] {}_vT_p$$

But by hypothesis $_vT_p$ may be *any* first order covariant tensor. We conclude, therefore, that

$$_vA_{mn} = \sum_a \sum_b {}_uA_{ab} \frac{\partial u_a}{\partial v_m} \frac{\partial u_b}{\partial v_n}$$

i.e. that $_vA_{mn}$ is a component of a second order covariant tensor.

§ 55. We may now summarize the tests by which it may be determined whether a given group of 4^n quantities is a tensor. (It is assumed, of course, that they are related, in some symmetrical way, to the four co-ordinate axes of some system.)

The group constitutes a tensor

(i) If each component in a given system is connected with all the components in another system in accordance with one of the laws of tensor-transformation.

(ii) If each component is the sum or difference of corresponding components of groups known to be tensors.

(iii) If the products of the components by the components of any arbitrarily chosen tensor are themselves components of a tensor.

CHAPTER XII

TENSOR-DIFFERENTIATION

§ 56. In the older dynamics and physics, which assumed a single universal space-time system, the laws of nature are very frequently expressed by linear differential equations, usually of the second order. In the theory of relativity, which admits an endless multiplicity of space-time systems, one may expect the same kind of thing to be true, with the difference that the differential equations must be *tensor*-equations, capable of preserving their essential features as they are transformed from any one system to another.

One naturally expects that differential tensor-equations would be built up by differentiating tensors. Unfortunately that anticipation is verified only in a roundabout way; for it is soon discovered that the differential coefficient of a tensor is not itself a tensor. For instance, let the tensor-component ${}_vT_m$ be differentiated with regard to the coordinate v_n; then we have

$${}_vT_m = \sum_a {}_uT_a \frac{\partial u_a}{\partial v_m}$$

whence

$$\frac{\partial}{\partial v_n}({}_vT_m) = \sum_a \frac{\partial}{\partial v_n}({}_uT_a)\frac{\partial u_a}{\partial v_m} + \sum_a {}_uT_a \frac{\partial^2 u_a}{\partial v_m \partial v_n}$$

$$= \sum_a \sum_b \left[\frac{\partial}{\partial u_b}({}_uT_a)\right]\frac{\partial u_a}{\partial v_m}\frac{\partial u_b}{\partial v_n} + \sum_a {}_uT_a \frac{\partial^2 u_a}{\partial v_m \partial v_n}$$

where (23 : 4) has been used in passing from the first line to the second. Now, if the second sum on the right were absent we could say that the differential coefficient of the first order tensor T_m was itself a second order tensor; but as things are that statement is evidently untrue. Thus we are driven to the conclusion that the ordinary process of differentiation when applied to a tensor introduces an element which is not universal but is characteristic only of the particular system in which the operation is carried out. If, then, we are to proceed further we must, as in previous analogous situations, seek some means of correcting the process of differentiation so that it may yield results of universal validity.

§ 57. The method we shall follow is Einstein's, slightly modified by Eddington. Since δv_m is a first order contravariant tensor and δs is invariant (that is a tensor of zero order), the differential coefficient dv_m/ds is also a first order contravariant tensor. If we multiply this by the first order covariant tensor T_m, the restricted product

$$\sum_m T_m \frac{dv_m}{ds}$$

is invariant by § 52. By "invariant" we mean, of course, that its value at the point-instant marked out by a given event-particle is independent of the system, so that

$$\sum_m \left({}_vT_m \frac{dv_m}{ds}\right) = \sum_a \left({}_uT_a \frac{du_a}{ds}\right) \qquad (57:1)$$

v_m and u_a being corresponding coordinates in the two

systems. Consider another near event-particle; then in virtue of (57 : 1)

$$\delta\left[\sum_m {}_v\left(T_m \frac{dv_m}{ds}\right)\right] = \delta\left[\sum_a {}_u\left(T_a \frac{du_a}{ds}\right)\right]$$

Divide by the invariant separation δs between the event-particles. Then we have in the limit

$$\frac{d}{ds}\left[\sum_m T_m \frac{dv_m}{ds}\right] = \sum_m \frac{dv_m}{ds}\frac{dT_m}{ds} + \sum_m T_m \frac{d^2v_m}{ds^2} \qquad (57:2)$$

is invariant. But by (23 : 4)

$$\frac{dT_m}{ds} = \sum_n \frac{\partial T_m}{\partial v_n}\frac{dv_n}{ds}$$

and it is evident that no difference will be made if we substitute p for m in the last term on the right of (57 : 2). Thus the invariant (57 : 2) may be written

$$\sum_m\sum_n \frac{\partial T_m}{\partial v_n}\frac{dv_m}{ds}\frac{dv_n}{ds} + \sum_p T_p \frac{d^2v_p}{ds^2} \qquad (57:3)$$

Now let the two event-particles lie upon a geodesic. Then by (32 : 15)

$$\frac{d^2v_p}{ds^2} = -\sum_m\sum_n \{mn, p\}\frac{dv_m}{ds}\frac{dv_n}{ds}$$

holds good for all systems. Making the substitution in (57 : 3), we have

$$\sum_m\sum_n \frac{\partial T_m}{\partial v_n}\frac{dv_m}{ds}\frac{dv_n}{ds} - \sum_m\sum_n\sum_p T_p\{mn, p\}\frac{dv_m}{ds}\frac{dv_n}{ds}$$

i.e.
$$\sum_m\sum_n \frac{dv_m}{ds}\frac{dv_n}{ds}\left[\frac{\partial T_m}{\partial v_n} - \sum_p T_p\{mn, p\}\right]$$

is invariant. Hence as in § 52 it follows that

$$\frac{\partial T_m}{\partial v_n} - \sum_p T_p \{mn, p\} \qquad (57:4)$$

is a second order covariant tensor. This, then, is the expression we are seeking; for it is based upon the differential coefficient of the tensor T_m and is yet a tensor.

It can be proved that in the case of a first order contravariant tensor, T^m, the corresponding expression is

$$\frac{\partial T^m}{\partial v_n} + \sum_p T^p \{mn, p\} \qquad (57:5)$$

We shall, however, need in the sequel only covariant tensors and their covariant tensor-derivatives.

§ 58. We proceed next to apply (57 : 5) in order to obtain a corresponding expression in the case of a covariant tensor of the second order.

If T_{mn} is the tensor, then by (50 : 1) we have

$$T_{mn} = A_m A'_n + B_m B'_n + C_m C'_n + D_m D'_n$$

where A_m, A'_n, etc., are components of suitably chosen first order tensors. For brevity let the right-hand side of this equality be written $\Sigma A_m A'_n$, and let T_{mn} be differentiated with regard to v_p. Then

$$\begin{aligned}\frac{\partial T_{mn}}{\partial v_p} &= \Sigma\left[\frac{\partial A_m}{\partial v_p} A'_n + A_m \frac{\partial A'_n}{\partial v_p}\right] \\ &= \Sigma\left[\left(\frac{\partial A_m}{dv_p} - \sum_q A_q \{mp, q\}\right) A'_n \right. \\ &\qquad\qquad \left. + \left(\frac{\partial A'_n}{\partial v_p} - \sum_q A'_q \{np, q\}\right) A_m\right] \\ &\quad + \sum_q \left[(\Sigma A_q A'_n)\{mp, q\} + (\Sigma A_m A'_q)\{np, q\}\right]\end{aligned} \qquad (58:1)$$

But $\Sigma A_q A'_n = T_{qn}$ and $\Sigma A_m A'_q = T_{mq}$. Hence (58 : 1) can be written

$$\frac{\partial T_{mn}}{\partial v_p} - \sum_q [T_{qn}\{mp, q\} + T_{mq}\{np, q\}]$$

$$= \Sigma\left[\left(\frac{\partial A_m}{\partial v_p} - \sum_q A_q\{mp, q\}\right)A'_n + \left(\frac{\partial A'_n}{\partial v_p} - \sum_q A'_q\{np, q\}\right)A_m\right]$$

Now each of the terms in the square bracket in the last line is a product of two factors, of which one is, by hypothesis, a first order covariant tensor, while the other was proved in § 57 to be a second order covariant tensor. The products are, therefore, components of tensors of the third order. Moreover, the coordinates (in the V-system) to which these third order components are related are in both cases the same: namely, v_m, v_n and v_p. It follows that, for a given pair of values of m and n (p is of course constant here), the sum within the square bracket is also a component of a third order tensor (§ 48). Finally, the same thing is true of the sum obtained by giving effect to the summation-sign before the bracket. We conclude, then, that

$$\frac{\partial T_{mn}}{\partial v_p} - \sum_q \Big[T_{qn}\{mp, q\} + T_{mq}\{np, q\}\Big] \quad (58:2)$$

is itself a component of a third order covariant tensor.

The corresponding result in the case of a second order contravariant tensor has a *plus* in the place of the *minus* in (58 : 2).

§ 59. We have already remarked that the differentiation of a tensor introduces an element which escapes from

the law of tensor-transformation and expresses a character of the differential coefficient that holds good only for a particular system. The nature of this non-universal element, in the case of covariant tensors, is indicated by the expressions in (57 : 4) and (58 : 2) which follow the *minus* sign. Note that in (57 : 4), where the differentiated tensor is of the first order, the non-universal element is related to two coordinate directions—namely, those of the tensor-component operated on (m) and of the axis parallel to which the differentiation takes place (n). In (58 : 2), where the tensor is of the second order, the non-universal character to be removed consists of two elements, which are related unsymmetrically to the directions specified for the tensor-component (m and n), while both are related in the same way to the direction of differentiation p. The sign of the non-universal characters is positive for covariant and negative for contravariant tensors.

If a coordinate system is Galilean, the g's are constants, and the three-index symbols of Christoffel all vanish (32 : 17). In that case the corrections needed to secure universality for the differential coefficients also vanish. In other words, the ordinary processes of differentiation applied to tensors produce coefficients which are themselves tensors for all Galilean systems.

§ 60. To find the correction needed to convert a second differential coefficient of T_m into a tensor-component, we have only to substitute (57 : 5), suitably modified, for T_{mn} in (58 : 2); for, as we have seen, (57 : 5) is a second order covariant tensor-component. To avoid confusion between the uses of the symbol p in the two

expressions, substitute r for it in (57 : 5); (58 : 2) then becomes

$$\begin{aligned}
&\frac{\partial^2 T_m}{\partial v_p \partial v_n} - \sum_r \left[T_r \frac{\partial}{\partial v_p} \{mn, r\} + \{mn, r\} \frac{\partial T_r}{dv_p} \right] \\
&\qquad - \sum_q \Bigg[\left(\frac{\partial T_q}{\partial v_n} - \sum_r T_r \{qn, r\} \right) \{mp, q\} \\
&\qquad\qquad + \left(\frac{\partial T_m}{\partial v_q} - \sum_r T_r \{mq, r\} \right) \{np, q\} \Bigg] \\
&= \frac{\partial^2 T_m}{\partial v_p \partial v_n} - \sum_r \frac{\partial T_r}{\partial v_p} \{mn, r\} - \sum_q \frac{\partial T_q}{\partial v_n} \{mp, q\} - \sum \frac{\partial T_m}{\partial v_q} \{np, q\} \\
&\qquad - \sum_r T_r \Bigg[\frac{\partial}{\partial v_p} \{mn, r\} \\
&\qquad\qquad - \sum_q [\{qn, r\} \{mp, q\} + \{mq, r\} \{np, q\}] \Bigg]
\end{aligned}$$

In the first line on the right-hand side of this identity no difference will be made on summation by substituting r for q. The required tensor-component then takes the tidier form

$$\begin{aligned}
&\frac{\partial^2 T_m}{\partial v_p \partial v_n} - \sum_r \left[\frac{\partial T_r}{\partial v_p} \{mn, r + \frac{\partial T_r}{\partial v_n} \{mp, r\} + \frac{\partial T_m}{\partial v_r} \{np, r\} \right] \\
&\qquad - \sum_r T_r \Bigg[\frac{\partial}{\partial v_p} \{mn, r\} - \sum_q [\{qn, r\} \{mp, q\} \\
&\qquad\qquad + \{mq, r\} \{np, q\}] \Bigg] \qquad (60:1)
\end{aligned}$$

As in § 58 this is a third order covariant tensor.

CHAPTER XIII

THE LAW OF GRAVITATION

§ 61. We are now in a position to understand how Einstein arrives at his law of gravitation. As we saw in § 56, it may be expected to consist of a set of differential equations of the second order having the tensor-character of validity in all coordinate systems. Moreover, in conformity with the principle that all movements in nature are determined by the intrinsic metrical properties of the region of space-time where they occur, the tensor-equations to be established must involve the space and time derivatives of the g's but no other variables.

The equations produced by equating to zero the third order covariant tensor-components deduced in (60 : 1) partly fulfil these conditions. They are linear differential equations and they contain second as well as first derivatives of the g's. The last point becomes clear when it is remembered that $\{mn, r\}$ is merely a condensed expression for

$$\sum_a \tfrac{1}{2} g^{ra} \left(\frac{\partial g_{ma}}{\partial v_n} + \frac{\partial g_{na}}{\partial v_m} - \frac{\partial g_{mn}}{\partial v_a} \right)$$

Hence $\dfrac{\partial}{\partial v_p} \{mn, \ r\}$ contains the second differential co-efficients, $\dfrac{\partial^2 g_{ma}}{\partial v_p \partial v_n}$, etc. On the other hand, (60 : 1) also involves an arbitrary tensor of which T_m and T_r are

components. If, therefore, it is to supply the equations we need, this superfluous element must be eliminated.

Now it will be noticed that in (60 : 1) n and p appear symmetrically in the first two terms but not in the third. This means that the tensor based upon

$$\frac{\partial}{\partial v_p}\left(\frac{\partial T_m}{\partial v_n}\right)$$

is not identical with the tensor based upon

$$\frac{\partial}{\partial v_n}\left(\frac{\partial T_m}{\partial v_p}\right)$$

If, then, (60 : 1) be rewritten with p and n interchanged, and if the resulting expression be subtracted from (60 : 1), we obtain

$$\sum_r T_r \Big[\frac{\partial}{\partial v_n}\{mp, r\} - \frac{\partial}{\partial v_p}\{mn, r\} + \sum_q [\{nq, r\}\{mp, q\} - \{pq, r\}\{mn, q\}]\Big] \qquad (61:1)$$

and since this is the difference between two covariant tensor-components of the third order it must itself be a tensor-component of that type and order. Moreover, its form shows that the tensor is the restricted product of an arbitrary first order covariant tensor T_r and the factor within the square brackets. It follows, by § 54, that this factor must be a component of a mixed tensor containing one contravariant index r and three covariant indices, m, n, p. The symbol q, which enters into the factor purely in connexion with the process of summation, is what Professor Eddington calls a "dummy" index.

Any other symbol could be substituted for it without change of significance; but the other four symbols indicate the coordinate directions involved in the component, and are therefore not dummies.

The mixed fourth order tensor of which the second factor of (61 : 1) is a component is the famous Riemann–Christoffel tensor, and is usually symbolized as B^r_{mnp}. Thus

$$B^r_{mnp} = \frac{\partial}{\partial v_n}\{mp, r\} - \frac{\partial}{\partial v_p}\{mn, r\} + \sum_q [\{nq, r\}\{mp, q\} - \{pq, r\}\{mn, q\}] \qquad (61:2)$$

§ 62. Now consider the tensor-equation

$$B^r_{mnp} = 0 \qquad (62:1)$$

Since the elements that enter into it are all either three-index symbols or their derivatives, it will evidently be satisfied by the g's in

$$\delta s^2 = \sum_m \sum_n g_{mn} \delta v_m \delta v_n$$

whenever those coefficients are constants. For the three-index symbols have the typical form shown in (32 : 17), and must plainly be zero if the g's are not functions of the coordinates. Thus we find that (62 : 1) is satisfied wherever space-time has the Galilean character expressed by the formula

$$\delta s^2 = -\delta r^2 - r^2\delta\theta^2 - r^2 \sin^2\theta\, \delta\phi^2 + c^2\delta t^2$$

But, by § 49, if the components of a tensor are zero in one system they are also zero in any other system to which they can be referred. It follows that if any expression

for δs^2 is such that it could, by a change of coordinates, be reduced to the Galilean form, the g's must satisfy (62 : 1). Moreover, it can be shown that no further condition is required. Thus $B^r_{mnp} = 0$ offers a necessary and sufficient condition for the absence of a permanent gravitational field.

§ 63. Where, as round the sun, a permanent gravitational field exists, the tensor-equations condensed into (62 : 1) cannot all be satisfied. It will be a useful exercise for the reader to verify this statement—assuming for the purpose that (36 : 4) expresses truly the metrical properties of the solar field. As an instance, take $m = 2$, $n = 1$, $p = 2$, $r = 1$, and determine for this case the value of (61 : 2) :

$$B^1_{212} = \frac{\partial}{\partial v_1}\{22, 1\} - \frac{\partial}{\partial v_2}\{21, 1\} + \sum_q[\{1q, 1\}\{22, q\} - \{2q, 1\}\{21, q\}] \tag{63:1}$$

The method of calculating the values of the three-index symbols has been explained in § 41. The reader will, therefore, have no difficulty in finding that

$$\{22, 1\} = -\tfrac{1}{2}g^{11}\frac{\partial g_{22}}{\partial r} = -(r-k);\ \{21, 1\} = \tfrac{1}{2}g^{11}\frac{\partial g_{22}}{\partial \theta} = 0,$$

whence

$$\frac{\partial}{\partial v_1}\{22, 1\} = -\frac{\partial}{\partial r}(r-k) = -1;\ \frac{\partial}{\partial v_2}\{21, 1\} = 0 \tag{63:2}$$

Proceeding to the second part of (61 : 2), it is necessary to evaluate the three-index symbols for $q = 1, 2, 3, 4$, in succession.

The results are as follows:

For $\{1q, 1\}$: $\{11, 1\} = \frac{1}{2} g^{11} \frac{\partial g_{11}}{\partial r}$ $\{12, 1\} = 0$

$\{13, 1\} = 0$ $\{14, 1\} = 0$

For $\{22, q\}$: $\{22, 1\} = -\frac{1}{2} g^{11} \frac{\partial g_{22}}{\partial r}$ $\{22, 2\} = 0$

$\{22, 3\} = 0$ $\{22, 4\} = 0$

For $\{2q, 1\}$: $\{21, 1\} = 0$ $\{22, 1\} = -\frac{1}{2} g^{11} \frac{\partial g_{22}}{\partial r}$

$\{23, 1\} = 0$ $\{24, 1\} = 0$

For $\{21, q\}$: $\{21, 1\} = 0$ $\{21, 2\} = 0$

$\{21, 3\} = 0$ $\{21, 4\} = 0$

Hence

$$\Sigma_q [\{1q, 1\} \{22, q\} - \{2q, 1\} \{21, q\}] = -\frac{1}{4} (g^{11})^2 \frac{\partial g_{11}}{\partial r} \frac{\partial g_{22}}{\partial r}$$

$$= \frac{1}{4} \left(1 - \frac{k}{r}\right)^2 \frac{k}{r^2} \left(1 - \frac{k}{r}\right)^{-2} . 2r = \frac{1}{2} \frac{k}{r} \qquad (63:3)$$

Adding the results of (63 : 2, 3) we obtain

$$B^1_{212} = -1 + \frac{1}{2} \frac{k}{r}$$

which proves that all the equations (62 : 1) are not satisfied by the coefficients of (32 : 17).

The reader may give himself further exercise by trying other sets of values for m, n and p. In some cases (62 : 1) will be satisfied, in other cases it will not.

§ 64. We have not yet reached the law of gravitation, but are well in sight of it. For the hypothesis we have assumed throughout our work * is that the special features which distinguish space-time around the sun gradually fade out with increasing distance until we come at length to a region of Galilean simplicity. It follows that the

* Einstein himself now inclines to a less simple hypothesis, but the argument here given is not materially affected.

law of gravitation must be closely related to $B^r_{mnp} = 0$. In fact, just as uniform motion is a special case of accelerated motion, so Galilean is a special case of gravitational space-time. The law of gravitation must, therefore, be a set of tensor-equations which includes (62 : 1) as a special case. The most obvious sets fulfilling this condition are those produced by restricting the Riemann–Christoffel tensor by making r equal to one of the covariant suffixes. For restriction will reduce the tensor from the fourth to the second order, with the result that the number of equations to be satisfied by the g's will be diminished. It is easily understood that when the rigour of the test is thus mitigated, peculiarities of space-time may survive which could not pass the challenge of the whole of the equations included in $B^r_{mnp} = 0$. On the other hand, the Galilean character which, as we have seen, passes the severer test, will also pass the less severe.

The only question remaining is: To which of the covariant suffixes is r to be equated? Let us begin by trying the effect of making $r = m$. Then the restricted tensor-component becomes (see § 51)

$$\begin{aligned}\sum_m B^m_{mnp} = \frac{\partial}{\partial v_n}\sum_m \{mp, m\} - \frac{\partial}{\partial v_p}\sum_m \{mn, m\} \\ + \sum_m\sum_q [\{nq, m\}\{mp, q\} - \{pq, m\}\{mn, q\}]\end{aligned} \tag{64 : 1}$$

Now since the double sum $\sum_m\sum_a \partial g_{pa}/\partial v_m$ is identical with $\sum_m\sum_a \partial g_{pm}/\partial v_a$ we have

$$\begin{aligned}\sum_m \{mp, m\} &= \sum_m\sum_a \tfrac{1}{2}g^{ma}\left(\frac{\partial g_{ma}}{\partial v_p} + \frac{\partial g_{pa}}{\partial v_m} - \frac{\partial g_{mp}}{\partial v_a}\right) \\ &= \sum_m\sum_a \tfrac{1}{2}g^{ma}\frac{\partial g_{ma}}{\partial v_p}\end{aligned} \tag{64 : 2}$$

Similarly $$\sum_m \{mn, m\} = \sum_m \sum_a \tfrac{1}{2} g^{ma} \frac{\partial g_{ma}}{\partial v_n}$$

Now it can be proved that in every case

$$\sum_m \sum_a g^{ma} \frac{\partial g_{ma}}{\partial v_p} = \frac{\partial}{\partial v_p} (\log g)$$

where g is the value of the determinant of the g's. The general proof involves a theorem about the differentiation of determinants with which it seems hardly worth while to burden an elementary treatise. It is, however, easy to give a proof for the special case in which the only g's which are not zero are g_{11}, g_{22}, g_{33} and g_{44}. For in that case the value of the determinant of the g's is the product

$$g_{11} \cdot g_{22} \cdot g_{33} \cdot g_{44} = g,$$

the minor of g_{11} is the product $g_{22} \cdot g_{33} \cdot g_{44}$, and the value of g^{11} is

$$g_{22} \cdot g_{33} \cdot g_{44}/g = 1/g_{11}$$

with corresponding values for g^{22}, etc. Thus:

$$g^{mm} \frac{\partial g_{mm}}{\partial v_p} = \frac{1}{g_{mm}} \frac{\partial g_{mm}}{\partial v_p} = \frac{\partial}{\partial v_p} (\log g_{mm})$$

and $$\sum_m \sum_m g^{mm} \frac{\partial g_{mm}}{\partial v_p} = \frac{\partial}{\partial v_p} (\log g_{11} + \log g_{22} + \text{etc.})$$
$$= \frac{\partial}{\partial v_p} (\log g)$$

Similarly

$$\sum_m \sum_m g^{mm} \frac{\partial g_{mm}}{\partial v_n} = \frac{\partial}{\partial v_n} (\log g)$$

In the simple case considered we should, then, have

$$\sum_m \{mp, m\} = \tfrac{1}{2} \frac{\partial}{\partial v_p} (\log g) \qquad \sum_m \{mn, m\} = \tfrac{1}{2} \frac{\partial}{\partial v_n} (\log g)$$

and it would follow that

$$\frac{\partial}{\partial v_n} \sum_m \{mp, m\} - \frac{\partial}{\partial v_p} \sum_m \{mn, m\} = 0$$

It will appear later that this special case is in fact the only one with which our theory has to deal, so that the cogency of the argument loses nothing by being confined to it. The result we have reached is, however, perfectly general.

Again, it is evident that on summation

$$\sum_m \sum_q [\{nq, m\}\{mp, q\} - \{pq, m\}\{mn, q\}] = 0$$

(For instance, the value when $m = 3, q = 2$ is cancelled by the value when $m = 2, q = 3$.)

We conclude that when $r = m$, the restricted (contracted) Riemann–Christoffel tensor vanishes identically. It is plain, therefore, that r must be equated, not with m, but either with n or with p. Now the interchange of n and p in (64 : 1) merely reverses the sign of the component; so it makes no difference which of those suffixes is chosen. Let us choose p. Then the tensor becomes

$$\begin{aligned}\sum_p B^p_{mnp} &= \frac{\partial}{\partial v_n}\sum_p \{mp, p\} - \frac{\partial}{\partial v_p}\sum_p \{mn, p\} \\ &\qquad + \sum_p \sum_q [\{nq, p\}\{mp, q\} - \{pq, p\}\{mn, q\}] \\ &= \sum_p \sum_a \frac{\partial}{\partial v_n}\left(\tfrac{1}{2} g^{pa} \frac{\partial g_{pa}}{\partial v_m}\right) - \frac{\partial}{\partial v_p}\sum_p \{mn, p\} \\ &\qquad - \sum_p \sum_q \sum_a \left(\tfrac{1}{2} g^{pa} \frac{\partial g_{pa}}{\partial v_q}\right)\{mn, q\} + \sum_p \sum_q \{nq, p\}\{mp, q\}\end{aligned} \tag{64:3}$$

Note that (64 : 2) has been used to simplify $\sum_p \{mp, p\}$ and $\sum_p \{pq, p\}$.

We have now reached the goal of our inquiry. Representing the second order tensor-component (64 : 3)

by the symbol G_{mn}, Einstein's Law of Gravitation takes the form

$$G_{mn} = 0 \qquad (64:4)$$

§ 65. It would be possible to use (64 : 4) to test the four coefficients in (36 : 4) and thus to establish or to destroy the credit of that formula. But although a useful exercise for the reader, this would not be a satisfactory logical procedure; for it could not prove that particular set of coefficients to be the only one admissible. It will be better, therefore, to take advantage of previous arguments only so far as to assume (as in § 36) that the separation between two near event-particles in the solar field is given by the formula

$$\delta s^2 = - A\,\delta r^2 - r^2\delta\theta^2 - r^2 \sin^2\theta\, \delta\phi^2 + Cc^2\delta t^2$$

which, in accordance with a familiar mathematical device, we shall use in the form

$$\delta s^2 = - e_{\alpha}\delta r^2 - r^2\delta\theta^2 - r^2 \sin^2\theta\, \delta\phi^2 + c^2e^{\beta}\delta t^2 \qquad (65:1)$$

e being the base of the natural logarithms.

Now the radial symmetry of the field shows (as in § 28) that α and β cannot be functions of θ or ϕ, and its temporal symmetry that they cannot involve t. They must, therefore, be functions of r only.

With the above assumptions, the determinant of the g's has the value $- c^2e^{\alpha+\beta}r^4 \sin^2\theta$, and the only values of g^{mn} which are not zero are:

$$g^{11} = - e^{-\alpha} \qquad\qquad g^{22} = - 1/r^2$$
$$g^{33} = - 1/(r^2 \sin^2\theta) \qquad\qquad g^{44} = e^{-\beta}/c^2$$

Remembering that $v_1 = r$, $v_2 = \theta$, $v_3 = \phi$, $v_4 = ct$, we have the following table of values:

m	n	p	$\partial g_{mn}/\partial v_p$	$\frac{1}{2}g^{mn}\partial g_{mn}/\partial v_p$
1	1	1	$-e^{a}\dfrac{da}{dr}$	$\frac{1}{2}\dfrac{da}{dr}$
2	2	1	$-2r$	$1/r$
3	3	1	$-2r\sin^2\theta$	$1/r$
3	3	2	$-r^2\sin 2\theta$	$\cot\theta$
4	4	1	$c^2e^{\beta}\dfrac{d\beta}{dr}$	$\frac{1}{2}\dfrac{d\beta}{dr}$

For all other combinations of m, n and p the results are zero. This table will be referred to as (65 : 2).

Our next task is to table the possible values of the four terms of (64 : 3).

(a) $\sum\limits_{p}\sum\limits_{a}\dfrac{\partial}{\partial v_n}\left(\frac{1}{2}g^{pa}\dfrac{\partial g_{pa}}{\partial v_m}\right)$ reduces to $\dfrac{\partial}{\partial v_n}\sum\limits_{p}\left(\frac{1}{2}g^{pp}\dfrac{\partial g_{pp}}{\partial v_m}\right)$, since g is zero unless $a = p$. Also by (65 : 2)

$$\sum_{p}\left(\tfrac{1}{2}g^{pp}\frac{\partial g_{pp}}{\partial v_m}\right) = \tfrac{1}{2}\frac{d}{dr}(a+\beta) + \frac{2}{r} \qquad [m = 1]$$
$$= \cot\theta \qquad [m = 2]$$

and is zero for the other values of m. Hence

$$\frac{\partial}{\partial v_n}\sum_{p}\left(\tfrac{1}{2}g^{pp}\frac{\partial g_{pp}}{\partial v_m}\right) = \tfrac{1}{2}\frac{d^2}{dr^2}(a+\beta) - \frac{2}{r^2} \qquad [m = n = 1]$$
$$= -\operatorname{cosec}^2\theta \qquad [m = n = 2]$$

and is zero for all other combinations of m and n.

$$(b)\ \sum_{p}\frac{\partial}{\partial v_p}\{mn, p\} = \sum_{p}\frac{\partial}{\partial v_p}\left[\tfrac{1}{2}g^{pp}\left(\frac{\partial g_{mp}}{\partial v_n} + \frac{\partial g_{np}}{\partial v_m} - \frac{\partial g_{mn}}{\partial v_p}\right)\right]$$

Now, of the 4^3 values of $\{mn, p\}$, only thirteen are not zero: namely,

$$\{11, 1\} = \tfrac{1}{2}\frac{da}{dr},\ \{12, 2\} = \frac{1}{r},\ \{13, 3\} = \frac{1}{r},\ \{14, 4\} = \tfrac{1}{2}\frac{d\beta}{dr}$$

$$\{22, 1\} = -re^{-a} \qquad \{23, 3\} = \cot\theta$$
$$\{33, 1\} = -re^{-a}\sin^2\theta \qquad \{33, 2\} = -\sin\theta\cos\theta$$
$$\{44, 1\} = \tfrac{1}{2}c^2e^{\beta-a}\frac{d\beta}{dr}$$

together with $\{21, 2\}$, $\{31, 3\}$, $\{32, 3\}$, and $\{41, 4\}$, whose values are the same respectively as those of $\{12, 2\}$, $\{13, 3\}$, $\{23, 3\}$, and $\{14, 4\}$.

When the foregoing values of $\{mn, p\}$ are differentiated with regard to the several values of v_p, the only derivatives which are not zero are the following.

(i) Derivatives with regard to v_1 ($= r$) :

$$\{11, 1\},\ \tfrac{1}{2}\frac{d^2\alpha}{dr^2} \qquad \{12, 2\},\ -\frac{1}{r^2} \qquad \{13, 3\},\ -\frac{1}{r^2} \qquad \{14, 4\},\ \tfrac{1}{2}\frac{d^2\beta}{dr^2}$$

$$\{21, 2\},\ -\frac{1}{r^2} \qquad \{22, 1\},\ -e^{-\alpha}\left(1 - r\frac{d\alpha}{dr}\right)$$

$$\{31, 3\},\ -\frac{1}{r^2} \qquad \{33, 1\},\ -e^{-\alpha}\sin^2\theta\left(1 - r\frac{d\alpha}{dr}\right)$$

$$\{41, 4\},\ \tfrac{1}{2}\frac{d^2\beta}{dr^2} \qquad \{44, 1\},\ \tfrac{1}{2}c^2e^{\beta-\alpha}\left\{\frac{d^2\beta}{dr^2} + \left(\frac{d\beta}{dr}\right)^2 - \frac{d\alpha}{dr}\frac{d\beta}{dr}\right\}$$

(ii) Derivatives with regard to v_2 ($= \theta$) :

$$\{23, 3\},\ -\operatorname{cosec}^2\theta \qquad \{32, 3\},\ -\operatorname{cosec}^2\theta$$

$$\{33, 1\},\ -re^{-\alpha}\sin 2\theta \qquad \{33, 2\},\ -\cos 2\theta$$

Hence $\sum\limits_p \dfrac{\partial}{\partial v_p}\{mn, p\}$ is zero except in the four following cases :

$$\tfrac{1}{2}\frac{d^2\alpha}{dr^2} \qquad [m = n = 1]$$

$$-e^{-\alpha}\left(1 - r\frac{d\alpha}{dr}\right) \qquad [m = n = 2]$$

$$-e^{-\alpha}\sin^2\theta\left(1 - r\frac{d\beta}{dr}\right) - \cos 2\theta \quad [m = n = 3]$$

$$\tfrac{1}{2}c^2e^{\beta-\alpha}\left\{\frac{d^2\beta}{dr^2} + \left(\frac{d\alpha}{dr}\right)^2 - \frac{d\alpha}{dr}\frac{d\beta}{dr}\right\} \qquad [m = n = 4]$$

(c) $\quad \sum\limits_p\sum\limits_q\sum\limits_a\left(\tfrac{1}{2}g^{pa}\dfrac{\partial g_{pa}}{\partial v_q}\right)\{mn, q\}$ reduces to

$$\sum_q\left[\sum_p \tfrac{1}{2}g^{pp}\frac{\partial g_{pp}}{\partial v_q}\right]\{mn, q\}$$

We have

$$\sum_q \left[\sum_p \tfrac{1}{2} g^{pp} \frac{\partial g_{pp}}{\partial v_q} \right] \{mn, q\} = \{mn, 1\} \left[\tfrac{1}{2} \frac{d}{dr} (\alpha + \beta) + \frac{2}{r} \right]$$

$$+ \{mn, 2\} \cot \theta$$

in accordance with the results obtained in (*a*). Hence, in accordance with the results obtained in (*b*), the only values of $\sum_p \sum_q \sum_a \left(\tfrac{1}{2} g^{pa} \frac{\partial g_{pa}}{dv_q} \right)$ which are not zero are:

$$\tfrac{1}{2} \frac{d\alpha}{dr} \left\{ \tfrac{1}{2} \frac{d}{dr} (\alpha + \beta) + \frac{2}{r} \right\} \qquad [m = n = 1]$$

$$\frac{1}{r} \cot \theta \qquad [m = 1, n = 2; \; m = 2, n = 1]$$

$$- re^{-\alpha} \left\{ \tfrac{1}{2} \frac{d}{dr} (\alpha + \beta) + \frac{2}{r} \right\} \qquad [m = n = 2]$$

$$- re^{-\alpha} \sin^2 \theta \left\{ \tfrac{1}{2} \frac{d}{dr} (\alpha + \beta) + \frac{2}{r} \right\} - \cos^2 \theta \qquad [m = n = 3]$$

$$\tfrac{1}{2} c^2 e^{\beta - \alpha} \frac{d\beta}{dr} \left\{ \tfrac{1}{2} \frac{d}{dr} (\alpha + \beta) + \frac{2}{r} \right\} \qquad [m = n = 4]$$

(*d*) Finally, the calculation of $\sum_p \sum_q \{nq, p\}\{mp, q\}$ involves only the following combinations which do not yield zero:

$$\{11, 1\}^2 + \{12, 2\}^2 + \{13, 3\}^2 + \{14, 4\}^2 \qquad [m = n = 1]$$

$$\{13, 3\} \{23, 3\} \qquad [m = 1, n = 2; \; m = 2, n = 1]$$

$$2 \{21, 2\} \{22, 1\} + \{23, 3\}^2 \qquad [m = n = 2]$$

$$2 \{31, 3\} \{33, 1\} + 2 \{32, 3\} \{33, 2\} \qquad [m = n = 3]$$

$$2 \{41, 4\} \{44, 1\} \qquad [m = n = 4]$$

Substituting the explicit values of the three-index symbols, we obtain the following table for

$$\sum_p \sum_q \{nq, p\} \{mp, q\}$$

$$\tfrac{1}{4}\left(\frac{d\alpha}{dr}\right)^2 + \frac{2}{r^2} + \tfrac{1}{4}\left(\frac{d\beta}{dr}\right)^2 \qquad [m = n = 1]$$

$$\frac{1}{r}\cot\theta \qquad [m = 1, n = 2;\ m = 2, n = 1]$$

$$-2e^{-\alpha} + \cot^2\theta \qquad [m = n = 2]$$

$$-2e^{-\alpha}\sin^2\theta - 2\cos^2\theta \qquad [m = n = 3]$$

$$\tfrac{1}{2}c^2 e^{\beta-\alpha}\left(\frac{d\beta}{dr}\right)^2 \qquad [m = n = 4]$$

We are now in a position to write out the expressions for G_{mn}, the values of the several items of the tensor-component (64 : 3) being taken from the tables just calculated.

$$G_{11} = \tfrac{1}{2}\frac{d^2}{dr^2}(\alpha+\beta) - \frac{2}{r^2} - \tfrac{1}{2}\frac{d^2\alpha}{dr^2} - \tfrac{1}{2}\frac{d\alpha}{dr}\left\{\tfrac{1}{2}\frac{d}{dr}(\alpha+\beta) + \frac{2}{r}\right\}$$

$$+ \tfrac{1}{4}\left(\frac{d\alpha}{dr}\right)^2 + \frac{2}{r^2} + \tfrac{1}{4}\left(\frac{d\beta}{dr}\right)^2$$

$$= \tfrac{1}{2}\frac{d^2\beta}{dr^2} + \tfrac{1}{4}\left(\frac{d\beta}{dr}\right)^2 - \tfrac{1}{4}\frac{d\alpha}{dr}\frac{d\beta}{dr} - \frac{1}{r}\frac{d\alpha}{dr}$$

$$G_{22} = -\operatorname{cosec}^2\theta + e^{-\alpha}\left(1 - r\frac{d\alpha}{dr}\right)$$

$$+ re^{-\alpha}\left\{\tfrac{1}{2}\frac{d}{dr}(\alpha+\beta) + \frac{2}{r}\right\} - 2e^{-\alpha} + \cot^2\theta$$

$$= e^{-\alpha}\left\{1 + \tfrac{1}{2}r\frac{d}{dr}(\beta-\alpha)\right\} - 1$$

$$G_{33} = e^{-\alpha}\sin^2\theta\left(1 - r\frac{d\alpha}{dr}\right) + \cos 2\theta$$

$$+ re^{-\alpha}\sin^2\theta\left\{\tfrac{1}{2}\frac{d}{dr}(\alpha+\beta) + \frac{2}{r}\right\} + \cos^2\theta$$

$$- 2e^{-\alpha}\sin^2\theta - 2\cos^2\theta$$

$$= \left[e^{-\alpha}\left\{1 + \tfrac{1}{2}\frac{d}{dr}(\beta-\alpha)\right\} - 1\right]\sin^2\theta$$

$$G_{44} = -\tfrac{1}{2}c^2e^{\beta-\alpha}\left\{\frac{d^2\beta}{dr^3} + \left(\frac{d\beta}{dr}\right)^2 - \frac{d\beta}{dr}\frac{d\alpha}{dr}\right\}$$

$$-\tfrac{1}{2}c^2e^{\beta-\alpha}\frac{d\beta}{dr}\left\{\tfrac{1}{2}\frac{d}{dr}(\alpha+\beta) + \frac{2}{r}\right\} + \tfrac{1}{2}c^2e^{\beta-\alpha}\left(\frac{d\beta}{dr}\right)^2$$

$$= -c^2e^{\beta-\alpha}\left\{\tfrac{1}{2}\frac{d^2\beta}{dr^2} + \tfrac{1}{4}\left(\frac{d\beta}{dr}\right)^2 - \tfrac{1}{4}\frac{d\alpha}{dr}\frac{d\beta}{dr} + \frac{1}{r}\frac{d\beta}{dr}\right\}$$

All other expressions for G_{mn} vanish.

§ 66. The final step is to determine the values of α and β by equating the foregoing expressions for G_{mn} to zero. It is to be observed that G_{22} and G_{33} yield the same equation, so that we have actually to deal only with three, not with four equations.

From $G_{44} = 0$ we obtain

$$\tfrac{1}{2}\frac{d^2\beta}{dr^2} + \tfrac{1}{4}\left(\frac{d\beta}{dr}\right)^2 - \tfrac{1}{4}\frac{d\alpha}{dr}\frac{d\beta}{dr} + \frac{1}{r}\frac{d\beta}{dr} = 0 \qquad (66:1)$$

and comparison of this result with $G_{11} = 0$ shows that

$$\frac{d\beta}{dr} = -\frac{d\alpha}{dr}$$

whence $$\beta = -\alpha + \text{constant} \qquad (66:2)$$

But, by our hypothesis, when r is infinite $e^{\alpha} = e^{\beta} = 1$; that is $\alpha = \beta = 0$. Hence the constant in (66 : 2) is zero and $\beta = -$.

Substituting in G_{22} (or in G_{33}) we have

$$e^{-\alpha}\left(1 - r\frac{d\alpha}{dr}\right) = 1$$

$$\frac{d}{dr}(re^{-\alpha}) = 1$$

$$re^{-\alpha} = r + k'$$

whence $$e^{\beta} = e^{-\alpha} = 1 + \frac{k'}{r}$$

and $$e^{\alpha} = \left(1 + \frac{k'}{r}\right)^{-1}$$

k' being the constant of integration.

Thus as the result of the whole investigation of the last four chapters we reach the conclusion that

$$ds^2 = -\left(1 + \frac{k'}{r}\right)^{-1} dr^2 - r^2\, d\theta^2 - r^2 \sin^2\theta\, d\phi^2 + c^2\left(1 + \frac{k'}{r}\right) dt^2 \qquad (66:3)$$

Now a repetition of the argument of § 36 would prove that $k' = -2GM/c^2$, the quantity which, in the formulæ of Chapters VIII, IX, was represented by $-k$. Substituting the old symbol for the new one, we return to the formula (36 : 4):

$$ds^2 = -\left(1 - \frac{k}{r}\right)^{1} dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta d\phi^2 + c^2\left(1 - \frac{k}{r}\right) dt^2$$

whose validity (with the momentous consequences depending on it) is thus established with complete universality.

Printed in Great Britain for the University of London Press, Ltd., by Hazell, Watson & Viney, Ld., London and Aylesbury.